The Role of Rubrics
in Advancing and Assessing Student Learning

Kenneth Wolf[†] and Ellen Stevens
University of Colorado at Denver and Health Sciences Center

Abstract

A rubric is a multi-purpose scoring guide for assessing student products and performances. This tool works in a number of different ways to advance student learning, and has great potential in particular for non-traditional, first generation, and minority students. In addition, rubrics improve teaching, contribute to sound assessment, and are an important source of information for program improvement. In this article, we discuss key features of a quality rubric, present an example of a rubric for assessing a social science research study, and describe three basic steps in designing an effective rubric.

Keywords: Rubrics, assessment, planning, instructional design.

While schoolteachers and their students have long seen the value of assessment rubrics, our experience in working with faculty is that rubrics have been largely ignored in higher education contexts (with the exception of Schools of Education). These multi-purpose scoring guides for assessing student products and performances work in a number of different ways to advance the goals of an educational program. Not only do rubrics contribute to student learning, they have great potential for non-traditional, first generation, and minority students. As well, rubrics improve teaching, provide feedback to students, contribute to sound assessment, and are an important source of information for program improvement.

So, what exactly are rubrics? How are they developed? What are their key features? Why are they useful? What are their limitations? What role can they play in program improvement? These questions, and more, will be addressed in this article.

Before we define and describe rubrics, here are a couple of scenarios to help set the stage (modified from Arter & McTighe, 2001, pp. x-xi):

> An undergraduate student in an American History course spent many hours working on her "museum display" on the Gold Rush. She received a "B" on her project with no other comments. She expressed concern that she had met the project guidelines and asked the professor what she could have done to get an "A." The professor responded, "I reserve 'A' for a highly creative project." When asked for

---

[†] Corresponding author's email: Kenneth.Wolf@cudenver.edu

> an example, the professor replied, "Well, you could have presented it from the point of view of the Native Americans affected by the Gold Rush."

What's the problem here…? There are no explicit performance criteria to inform students in creating their projects or to guide the professor in assessing them. A rubric here could help address this situation.

How do you think this student felt? Probably the same way that students in any course feel when the criteria for an assignment are ambiguous and the assessment seems arbitrary. When the curriculum is "hidden," students who can't guess what the expectations are will be more at risk than those who know how to "play the game" (Jackson, 1990). A good rubric can take the mystery out of assignments for *all* students. As Eisner notes: "More than what educators say, more than what they write in curriculum guides, evaluation practices tell both students and teachers what counts. How these practices are employed, what they address and what they neglect, and the form in which they occur speak forcefully to students about what adults believe is important" (Eisner, 1991, p. 81).

Now, let's look at another scenario:

> In an English department class, a professor introduced her students to the qualities of an effective oral presentation by showing them videotaped examples of excellent, as well as poor, speeches and presentations. Guided by the teacher, the students identified four key criteria (traits) that they agreed were important for an effective speech—content, organization, delivery, and language. They defined each of these and what would constitute strong, middle, and weak performance on each trait. They then referred to these performance criteria when preparing their own speeches, and the teacher used the same criteria when providing feedback on, and grading, their presentations.

What's going on in this scenario? Not only are there criteria that define the features of a speech, but the professor has shown strong and weak examples of oral presentations and even invited the students to generate evaluation criteria based on these examples and their own experiences. Clearly, both students and professor use the criteria in talking about and giving feedback on the speeches. In other words, the learning process is anchored by a rubric--a scoring tool used to evaluate a performance in a given outcome area based on a list of criteria describing the characteristics of products or performances at varying levels of accomplishment.

A Rubric for Springboard Diving

We always have criteria in mind when we evaluate something–whether it's a piece of art or a dive off a springboard. It's just that these criteria aren't always explicit, sometimes even to ourselves. When we judge a springboard diver's performance as good or bad, for example, we are basing that judgment on something. We have some criteria in mind. Maybe it's the number of body rotations or the splash the diver makes on entry. Maybe it's something that really has nothing to do with the performance itself such as the diver's

smile or nationality.

As we become more informed about springboard diving, though, we may begin to draw on the five criteria used by the professional association (Federation Internationale de Natation, 2006): Starting Position, Take Off, Approach, Flight, and Entry. These criteria are then elaborated in a rubric that describes what we mean by each. "Entry," for example, is based on a number of considerations about body position. "The entry into the water shall in all cases be vertical, or nearly so, with the body straight, the feet together and the toes pointed. When the entry is short or over, the judge shall deduct according to his opinion (p. x)." Each of these criteria is then described on six levels of performance from "complete failure" to "very good" (see Table 1).

A rubric in springboard diving makes it more clear to the judges how to rate the performance, though these judges still need to draw on their extensive professional knowledge in applying these criteria. As well, coaches study the criteria so that they can provide effective instruction to their athletes. And the athletes themselves examine the criteria to guide them in planning and perfecting their dives. In the same fashion, for an assignment in a course or for other types of learning experience, such as studios or internships, learning is best achieved if all participants are clear about the criteria for the performance and the levels at which it will be assessed.

Table 1. Springboard Diving Rubric

	Complete Failure	*Unsatisfactory*	*Deficient*	*Satisfactory*	*Good*	*Very Good*
Starting						
Take-off						
Approach						
Flight						
Entry						

Developing a Rubric

Sometimes rubric development stops after the performance criteria have been identified and performance levels established (as portrayed in Table 1), but the best rubrics include another step in which each of the cells in the matrix contains a description of the performance at that level. These three steps in designing a rubric will be discussed in the following section, though the particulars can vary across rubrics depending upon, for example, the context for the course or nature of the learning being assessed..

Identifying Performance Criteria. The first step in developing a rubric is to identify the criteria that define the performance. Suppose the performance task or expected learning outcome is that "students will be able to give an effective oral presentation." What are the key features or criteria of an effective oral presentation? While it depends upon the purpose and context for the speech, four general criteria might be identified: delivery, content, organization, and physicality.

Three to six criteria seem to work best. It is not so many that it overwhelms the memory and not so few that meaningful distinctions in the performance can't be made. Sometimes these criteria can be weighted as well. There may be one or two criteria that are valued more than the others and they could be given a higher value when calculating the overall score for the performance or product.

Another important consideration is that the performance to be assessed should be observable and measurable. Some descriptions of learning outcomes or performance criteria are so vague that accurate measurement is difficult. For example, if the criterion is that "Students will know the states of the union," it may not be clear what "know" means. Does 'knowing" mean that students need only to be able to list the states, or be able to fill in the names on a map, or draw a map of the United States, or discuss the history of the state, or ….? The measurement problem can be lessen if the performance to be assessed is described with more specific action verbs where possible, such as list, identify draw, discuss, explain, compare, critique, predict, and so on.

Often the performance criteria are determined ahead of time by the instructor or a professional organization, but sometimes they can be created by the students in a course, especially if the assignment is new to the instructor. Having students generate the criteria for assessing the performance can serve several purposes. Engaging students in a discussion about "What makes for a good speech" (or essay or model or dance or…) can help them deepen and internalize their understanding of the criteria for a quality performance in that particular area. As well, involving students in this conversation before they begin the assignment or project can help them make more informed choices as they begin to identify the topic for their laboratory study, the medium for their performance, or the design for their model. Another benefit is that students can sometimes offer insights into the performance that the instructor may not have envisioned. When a student asks if their oral presentations can be a video of themselves before a live audience rather than a live in person in class presentation, it can open possibilities the instructor hadn't considered. An additional pedagogical benefit is that the students' comments can reveal to the instructor misconceptions that students may have about the topic, and the instructor can adjust his or her teaching of these concepts accordingly. A valuable activity can be to make a list of the assessment criteria that students identify as the project is introduced and another list again after they have completed the project, and then have them compare their pre-and-post lists to see if their understanding of the key concepts have changed or deepened. Even if the rubric has already been developed in advance however, asking students to engage in a discussion about the assessment criteria before the rubric is handed out can still be a valuable activity for many of these same reasons.

Setting Performance Levels. The second step in the process is to decide how many levels of performance are appropriate for the assessment. Typically, rubrics have from three to six rating levels. What drives the choice of the number of levels is the purpose for the assessment. If the main purpose is to make summative decisions, such as whether someone will pass or fail a course or an exam for example, then fewer levels are better. The fewer the levels of performance for the rater to consider, the greater the reliability and effi-

ciency in scoring the performance. The more levels, the lower the reliability in scoring and the more time it will take for raters to make the decision.

If, however, the primary purpose of the assessment is formative, or to give feedback to learners to support them in improving their performance, then more performance levels (and more performance criteria) give the learner more specific information about the features of the performance that need attention. The trade-off again is that the greater number of scoring levels and performance criteria, the more time it takes the rater to assess the performance.

The headings for the different performance levels can vary depending upon the purpose and contexts for the assessment. For some contexts, developmental language is the best choice, such as "Emerging, Developing, Arrived." A developmental scale is respectful to the learner and recognizes that all of us are learners in any number of areas. The emphasis is on growth. Other times, more mastery-oriented language is appropriate as in "Below Proficient, Proficient, Above Proficient." If the purpose of the assessment is to demonstrate whether or not students have met the standards for the course or program or profession, then identifying whether a learner is proficient or not is the key. Sometimes, numbers are used instead of words, while at other times numbers and words are used together (see Table 2).

Table 2. Performance Criteria and Levels for Speech Rubric

	Below Proficient (1)	*Proficient (2)*	*Beyond Proficient (3)*
Delivery			
Content			
Organization			
Physicality			

Creating Performance Descriptions. The third step in the process is to write a description for each cell in the matrix. For example, "delivery" is described in a brief statement at each of the three performance levels (see Table 3). The challenge in creating these paragraphs is to provide enough information to guide the creation and scoring of the project, but not so much that it overwhelms the reader or the performer. Keep in mind that the rubric is not intended to replace the instructor but instead to guide and support him or her in exercising informed judgment.

Parallel structure across descriptions for each criterion (e.g., delivery) is important. The more parallel the descriptions are in form and content, the more dependable and efficient the scoring will be. One way to increase parallelism across descriptions is to identify a set of attributes for each criterion and then build each statement around those attributes. For example, the "delivery" descriptions were developed around three attributes: volume, pacing, and rapport. The same process is then followed for the other three criteria (e.g. content, organization, physicality) until all of the cells in the matrix are completed (see Table 4).

Table 3. Speech Rubric with Performance Statements for the "Delivery" Criterion

	Below Proficient (1)	*Proficient (2)*	*Beyond Proficient (3)*
Delivery • Volume • Pacing • Rapport	It is difficult to hear the speaker, and the pace is either too slow or too fast. Speaker has little connection with audience.	Speaker is easy to hear and pace keeps audience's attention.	Speaker varies volume to fit the message, with a pace that is appropriate to the rhythms of the topic. Audience is clearly engaged.

When using the rubric in making an overall decision about a performance, the final rating can be based on an analytic process of adding up the scores for each of the four criteria (i.e., content, delivery, language, physicality) and calculating an average, or, alternatively, by looking over the ratings for the four criteria and making a holistic judgment that considers each of the scores but blends them in an overall judgment-based rating process. For example, if the scores were delivery = 2, content = 3, organization = 2, and physicality = 3, then an analytical rating (assuming equal weighting of the four criteria) would give an overall mean score of 2.5. A holistic rating might end up as a 2 or 3 however, depending upon the rater's overall sense of the performance. When the criteria are not equally weighted, numerical calculations need to be adjusted accordingly

Rubric for Assessing a Social Science Research Study

The rubric presented in this section was developed by Kenneth Wolf (a co-author of this article) and his colleagues in the School of Education and Human Development for use in research methods classes for students who are earning a master's degree in education or counseling (see Table 5). The main assignment for the course, which counts for half of the course grade, is for students to work together in small groups to design and carry out small-scale research studies on topics in their fields. Students are encouraged to conduct studies that advance the learning or development of their students or clients, or that contribute in some way to the organizations in which they work. Students in education might, for example, conduct a pilot experimental study to examine the effectiveness of a new literacy curriculum that their school is considering purchasing, interview Latino parents about their understanding of the school district's policies on bilingual education, or observe an individual student on the playground as a way of understanding that student's social skills and development. Students first submit a research proposal and receive ungraded, written feedback (based on the rubric). At the end of the semester they turn in their completed studies and receive written feedback along with an overall rating based on the rubric performance levels (e.g., proficient).

Table 4. Speech Rubric

	Below Proficient (1)	*Proficient (2)*	*Beyond Proficient (3)*
Delivery • Volume • Pacing • Rapport	It is difficult to hear the speaker, and the pace is either too slow or too fast. Speaker has little connection with audience.	Speaker is easy to hear and pace keeps audience's attention.	Speaker varies volume to fit the message, with a pace that is appropriate to the rhythms of the topic. Audience is clearly engaged.
Content • Accuracy • Relevance • Organization	The content of the speech is inaccurate or incomplete, or not relevant to the topic or audience. The sequence of ideas is confusing.	The content is accurate and complete, and relevant to topic and audience. The content is well sequenced and the relationship among ideas clear.	The content is precise and comprehensive, and customized to the audience and appropriate for the topic. The sequence and organization of ideas are powerful.
Language • Vocabulary • Enunciation • Grammar	Vocabulary is simplistic or trite, or is not appropriate to audience or topic. Speech is sprinkled with "ums" or is difficult to understand. Speaker makes many grammatical mistakes.	Vocabulary is appropriate to audience and topic. Speech is clear and easy to understand. Grammar and syntax are sound.	Vocabulary is rich and vivid, and appropriate to audience and topic. Speech is clear and easy to understand, with careful attention to pronunciation. Grammatical and syntactical structures are complex and effective.
Physicality • Body Movement • Eye Contact • Facial Expression	Body movement is too much or too little. Speaker displays little eye contact and facial expression.	Body movement is appropriate to the context. Speaker makes regular eye contact with audience and varies facial expressions.	Speaker customizes body movement and gestures to context and topic. Speaker engages audience through varied and compelling eye contact and facial expressions.

Table 5. Rubric for Research Project in Education

	Below Proficient	*Proficient*	*Above Proficient*
Abstract	The <u>abstract</u> is missing, incomplete, or inaccurate.	The <u>abstract</u> summarizes the study in 50-150 words (essentially drawing a sentence from each of the main sections of the completed research report).	The <u>abstract</u> concisely summarizes the study in 50-150 words.
Introduction	The <u>introduction</u> section may be incomplete or unclear. Potential problems may include a vague problem statement, research question(s) may not be measurable, or constructs may not be clearly defined.	The <u>introduction</u> section includes a rationale, problem statement, literature references and research question(s). The rationale and problem statement are clear and credible. Three or more literature references are cited. The research question is stated and can be addressed with empirical evidence. Constructs are defined and variables explained.	The <u>introduction</u> section is complete and clear. Additionally, the rationale and problem statement are compelling (and may be linked to a conceptual framework) and the research question(s) insightful.
Methods	The <u>methods</u> section may be incomplete or unclear. Possible problems may include insufficient information about subjects/informants, instruments not fully described in terms of their conceptualization or aligned with the research questions, or procedures not accurately reported.	The <u>methods</u> section provides essential information about the subjects, data collection procedures, and, if appropriate, treatment. The research question has been translated into appropriate choices at the design level. Subjects are described in terms of number and important characteristics. Data sources and collection procedures are described in terms of underlying conceptualizations. If appropriate, scales are described, and examples of items given. Data collection protocols (e.g., questionnaires, interview questions, structured observation protocols) are included in the appendix.	The <u>methods</u> section provides essential information about the subjects, data collection procedures, instruments, procedures, and, if appropriate, treatment. In addition, the instrument or procedures, for example, might represent a novel and insightful approach to the research problem.

Table 5. Rubric for Research Project in Education (cont'd.)

	Below Proficient	*Proficient*	*Above Proficient*
Results	Results are inaccurate or incompletely presented. Typical problems include incorrect statistical analyses in quantitative studies and unsupported claims in qualitative-type studies.	The results section in a quantitative study presents only the "facts." Brief and accurate interpretation is offered, indicating understanding of how the data respond to the research questions. Tables or graphs are easy to interpret and correctly present the data. In a qualitative study, results and interpretation may be interwoven, and each theme is illustrated with two or more data segments (e.g., quotes from informants).	Results are correctly presented and the analyses are extensive and sophisticated.
Discussion/ Conclusion	The discussion section may be incomplete or not clearly connected to the results.	The discussion section soundly interprets the findings. The discussion section may also include conclusions, limitations of the study, recommendations for action, and future study directions.	The discussion section soundly interprets the findings and is carefully connected with all sections of the report, including the introduction, research questions, instruments, and results.
Limitations	Limitations of the study are not discussed.	Limitations of the study are discussed.	Limitations are extensively described.
References	References may be missing, incomplete, or incorrectly cited.	References are given (and correctly cited in the body of the report and included on a separate reference page in APA format).	References are correctly cited in body of the report and on a separate reference page in APA format.
Written Report	The written report is incomplete or unclear. Typical problems include missing or inadequately described sections.	The written report is clear and well organized. The vocabulary in the report demonstrates an understanding of key terms and concepts (e.g., construct, subject, treatment). The report contains few mechanical errors (e.g., punctuation) and is in APA format. Study is ethical.	The written report is clear and well organized and demonstrates an understanding of basic and advanced research concepts and terms.

In this course, both "beyond proficient" and "proficient" translate into "A" grades on the projects, but "beyond proficient" recognizes performances that go beyond what was required, which is not an uncommon occurrence with graduate students who may be presenting their findings at their school's faculty meetings or to school boards. "Below proficient" performances most often result in "B" grades since graduate students' projects typically suffer from only minor omissions or problems. The "beyond proficient" rating assumes that the students have demonstrated all of the features of a "proficient" performance, but with remarkable grace or insight.

A group's overall rating for their project could be based on an analytical averaging of the ratings for each of the individual sections of the report (e.g., "proficient" for the abstract, "beyond proficient" for the introduction section), assuming each of the sections is of equal weight or value. However, these projects are rated in a more holistic manner in which the faculty member considers a group's ratings on the individual sections but then steps back and makes an overall rating (while keeping each of the individual section ratings in mind), recognizing that sometimes the whole is larger than its parts.

Benefits of Rubrics

Rubrics contribute to student learning and program improvement in a number of ways—some obvious, others less so.

Rubrics make the learning target more clear. If students know what the learning target is, they are better able to hit it (Stiggins, 2001). When giving students a complex task to complete, such as a building an architectural model or putting together a portfolio of their best photographs, students who know in advance what the criteria are for assessing their performance will be better able to construct models or select photographs that demonstrate their skills in those areas.

Rubrics guide instructional design and delivery. When teachers have carefully articulated their expectations for student learning in the form of a rubric, they are better able to keep the key learning targets front and center as they choose instructional approaches and design learning environments that enable students to achieve these outcomes (Arter & McTigue, 2001).

Rubrics make the assessment process more accurate and fair. By referring to a common rubric in reviewing each student product or performance, a teacher is more likely to be consistent in his or her judgments. A rubric helps to anchor judgments because it continually draws the reviewer's attention to each of the key criteria so that the teacher is less likely to vary her application of the criteria from student to student. Furthermore, when there are multiple raters (e.g., large lecture classes that use teaching assistants as graders), the consistency across these raters is likely to be higher when they are all drawing on the same detailed performance criteria. Additionally, a more prosaic benefit is the decided decrease in student complaints about grades at semester's end.

Rubrics provide students with a tool for self-assessment and peer feedback. When stu-

dents have the assessment criteria in hand as they are completing a task, they are better able to critique their own performances (Hafner & Hafner, 2004). A hallmark of a professional is the ability to accurately and insightfully assess one's own work. In addition, rubrics can also be used by classmates to give each other specific feedback on their performances. (For both psychometric and pedagogical reasons, we recommend that peers give only formative feedback that is used to help the learner make improvements in the product or performance, and not give ratings that are factored into a student's grade.)

Rubrics have the potential to advance the learning of students of color, first generation students, and those from non-traditional settings. An often unrecognized benefit of rubrics is that they can make learning expectations or assumptions about the tasks themselves more explicit (Andrade & Ying, 2005). In academic environments we often operate on unstated cultural assumptions about the expectations for student performance and behavior and presume that all students share those same understandings. However, research by Lisa Delpit (1988) and Shirley Heath (1983), for example, highlights the many ways that expectations in schools are communicated through subtle and sometimes unrecognizable ways for students of color or non-native English speakers who may have been raised with a different (but valid) set of rules and assumptions about language, communication, and school performance itself.

Limitations of Rubrics

While well-designed rubrics make the assessment process more valid and reliable, their real value lies in advancing the teaching and learning process. But having a rubric doesn't necessarily mean that the evaluation task is simple or clear-cut. The best rubrics allow evaluators and teachers to draw on their professional knowledge and to use that professional knowledge in ways that the rating process doesn't fall victim to personality variations or limitations of human information processing.

A serious concern with rubrics, however, is how long it takes to create them, especially writing the descriptions of performances at each level. With that in mind, rubrics should be developed for only the most important and complex assignments. Creating a rubric that is used to determine whether students can name the parts of speech would be like using a scalpel to cut down a tree: Good instrument, wrong application.

Another challenge with rubrics is that if poorly designed they can actually diminish the learning process. Rubrics can act as a straitjacket, preventing creations other than those envisioned by the rubric-maker from unfolding. ("If it is not on the rubric, it must not be important or possible.") The challenge then is to create a rubric that makes clear what is valued in the performance or product—without constraining or diminishing them. On the other hand, the problem with having no rubric, or one that is so broad that it is meaningless, is to risk having an evaluation process that is based on individual whimsy or worse—unrecognized prejudices. Though not as dangerous as Ulysses' task of steering his ship between the two fabled monsters of Greek mythology, Scylla and Charybdis, a rubric-maker faces a similar challenge in trying to design a rubric that is neither too narrow nor too broad.

While not a panacea, the benefits of rubrics are many—they can advance student learning, support instruction, strengthen assessment, and improve program quality.

References

Andrade, H., & Ying, D. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, *10*(3), 1-11.

Arter, J. & McTighe, J. (2001). *Scoring Rubrics in the Classroom*. Thousand Oaks, CA: Corwin Press.

Delpit. L. (1988). The Silenced Dialogue: Power and Pedagogy in Educating Other People's Children. *Harvard Educational Review*, *58*(3), 280-298.

Eisner, E. (1991). *The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice*, New York: Macmillan.

Federation Internationale de Natation. (2006). *Rules and Regulations: FINA Diving Rules 2005-2009*. Retrieved January 27, 2006 from http://www.fina.org.

Hafner, J. C., & Hafner, P. M. (2004). 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.

Heath, S. B. (1983). *Ways with Words.* Cambridge: Cambridge University Press.

Jackson, P. (1990*). Life in Classrooms.* New York: Teachers College Press.

Stiggins, R. (2001). *Student-Involved Classroom Assessment* (3rd ed.). New York: Merrill.