

Developing Scoring Rubrics for ESL Writing Assessment

Anthony Del Vecchio*

Keywords: rubric, scoring, assessment, criteria, writing

Abstract

This paper examines the use of scoring rubrics to assess ESL learners' writing proficiency. Two broad categories of rubrics are defined: a general criteria model and a specific criteria model. The specific criteria model is further delineated into two categories, referred to herein as band assessment and paradigm assessment. Examples of the various rubrics are presented and an explanation as to how they were developed follows. The applicability of the different rubric types is then examined with regard to the type of writing involved and the rater's assessment objectives. Specific issues related to the usefulness and reliability of writing assessment are also considered as pertains to objectivity, consistency, and student feedback.

Introduction

Asked to assess a student's piece of writing and to assign a letter grade from A+ to F without any specific direction, three different teachers are apt to arrive at three entirely different grades. What is more worrisome is that one teacher may give the same student two different grades on two different days for the exact same paper, depending on a variety of factors such as busy schedules, heavy workloads, outside distractions, and so on. As Carnegie Mellon (n.d.) notes, "Grading consistency is difficult to maintain over time because of fatigue, shifting standards based on prior experience, or intrusion of other criteria" (Advantages of Using Rubrics, para. 1).

Furthermore, unlike mathematics or science, writing is an art, and while subject to some generally agreed-upon rules and conventions, there may exist a difference of opinion among teaching professionals as to what constitutes "good writing." Thus, writing assessment can often be highly subjective and, consequently, unfair to students. A common framework to guide writing assessment, particularly within schools and universities with unified curriculums, can reduce to some degree this subjectivity and assessment variability.

* 非常勤講師／第2外国語

One solution to issues of consistency and subjectivity in writing assessment lies with rubrics, which are an attempt to define a set of criteria that properly represent the degree to which a piece of writing performs a particular set of stylistic and/or communicative functions. Rubrics act as a set of guidelines that work to reduce variability both from the individual teacher's perspective and from the perspective of intra-departmental variability, e.g., across teachers using a unified curriculum and/or a common set of teaching goals and objectives.

A rubric can be defined as a “scoring tool that explicitly represents the performance expectations for an assignment or piece of work. A rubric divides the assigned work into component parts and provides clear descriptions of the characteristics of the work associated with each component, at varying levels of mastery” (Carnegie Mellon, n.d., What are rubrics?, para 1). That is, there are two features common to all rubrics: “(1) a list of criteria, or “what counts” in a project or assignment; and (2) gradations of quality, with descriptions of strong, middling, and problematic work” (Andrade, 2000, p. 13).

Why Use Scoring Rubrics?

The quality and consistency of student writing assessment can be enhanced through the discipline imposed by using scoring rubrics. Rubrics take much of the “guesswork” out of both defining the educator's course goals at the onset and ultimately in the assessment of students' achievement of those goals as represented by the finished writing product. Andrade (2005) explains that rubrics are useful tools for educators in that they help to:

...clarify our learning goals, design instruction that addresses those goals, communicate the goals to students, guide our feedback on students' progress toward the goals, and judge the final products in terms of the degree to which the goals were met. (p. 27)

Moreover, (Carnegie Mellon, n.d.) list four advantages of scoring rubrics from the instructor's perspective:

1. Use of an “explicit and descriptive set of criteria” helps ensure long-term consistency of an instructor's grading standards.
2. Grading time can be reduced by reducing uncertainty and by referring to rubric descriptions associated with a score.
3. Rubrics can help ensure consistency and reduce systematic bias when multiple graders are involved.
4. Rubrics can help the instructor identify skill areas and concepts that are in need of improvement.

Types of Scoring Rubrics

For purposes of this paper I have divided scoring rubrics into two main groups: general criteria rubrics and specific criteria rubrics. With the general criteria model, a single numerical score, letter grade, or proficiency description is assigned to a piece of writing based on the overall quality of the work. In other words, writing is assessed holistically and “not judged according to individual mistakes that [writers] make but by the overall effectiveness of [their] writing” (Rogers, p. 573). With the specific criteria model, individual criteria such as content, organization, grammar, mechanics, etc., are established which are then quantitatively or qualitatively assessed in turn for merit on a numerical scale.

General Criteria Rubrics

General criteria rubrics can represent a great savings in time when raters are required to judge large numbers of writing samples, yet they still maintain a reasonable degree of consistency in writing assessment. Thus, they are particularly appropriate for scoring short answer or long answer questions on standardized tests or student examinations in the classroom. General criteria rubrics work best when a relatively narrow range of scores is desired and when a more general description of writing ability is sufficient for the rater’s purposes.

A good example of a general criteria rubric is the type used in connection with the TOEFL (Test of English as a Foreign Language), an examination that includes two writing sections: an Integrated Writing Task, which requires test takers to summarize and compare reading passages and lectures, and an Independent Writing Task, in which test takers express an opinion or personal preference (Rogers, 2007, p. 560). Both tasks are rated holistically for content and linguistic skill on a scale of zero to five, with five being the highest possible score. Rogers (2007) presents examples of rubrics for both tasks that are similar to those used on the TOEFL test. For example, for the Integrated Writing Task, writers can attain the highest possible score of five if the writing conforms to the following criteria:

5. Includes all of the important information from the passage and accurately relates it to the important information in the lecture. The response is clearly organized. A variety of sentence structures are used and sophisticated vocabulary is employed. There may be infrequent grammatical and mechanical errors. These mistakes will not make it difficult to understand the ideas and relationships the writer describes. (p. 573)

Similarly, writers can achieve the highest possible score on the Independent Writing Task if their writing conforms to the highest standards expressed in that rubric:

5. Strongly indicates the ability to write a well-organized, well-developed, and logical response. Specific examples and details support the main ideas. All of the elements of the response are unified and cohesive. A variety of sentence structures are used successfully, and sophisticated

vocabulary is employed. Grammatical and mechanical errors are infrequent but a few minor mistakes may occur. (p. 650)

Figure 1 is an example of a general criteria rubric that I designed for assessing elementary paragraph writing skill for ESL (English as a Second Language) learners. In that rubric, I attempted to define at five different levels how a paragraph might be evaluated. In particular, descriptions all include such standard paragraph elements as organization, unity, coherence, sentence complexity, grammar, and vocabulary. The wording in the descriptions in the different levels all maintain generally parallel syntax and use approximately the same number of words in order to maintain a high degree of consistency within the rubric itself. An attempt was made in all descriptions to state as clearly and concisely as possible what general criteria represents writing ability at each level. Using this rubric, then, a student would achieve the highest possible rating (five out of five) if the paragraph conforms to the following general criteria:

Very well-organized writing with a logical order of presentation. The main idea is clearly expressed in a topic sentence and supported by sentences using specific examples, explanations, reasons, etc. Paragraph shows unity throughout with good use of transition words. A variety of sentence types are used, including simple, compound, and complex sentences. Higher-level vocabulary is used. Few, if any, grammatical or mechanical errors appear.

Conversely, a student would achieve the lowest possible rating (one out of five) if the writing conforms to the following general criteria:

Writing not in basic paragraph form. Shows little to no attempt at organization. Consists mainly of a random series of incoherent or irrelevant sentences that do not support a main idea. Sentences may be repetitive or so vague as to lack any real meaning. Rudimentary vocabulary is primarily used. Serious spelling and/or grammatical errors and lack of transition words make it very difficult to understand any intended message.

As mentioned earlier, general criteria rubrics can be useful tools for quickly assessing large numbers of writing samples in which only a score is required, i.e., when feedback to the writer is not intended. From the student's point of view, however, general criteria rubrics are not appropriate as instructional tools because they are, by necessity, overly general. (Brown University, 2015) notes several additional issues related to the use of general criteria rubrics with respect to their use as instructional tools, in particular for ELLs [English Language Learners]:

...raters can assign too much weight to grammatical forms and conventions that take years for ELLs to master. This approach also does not provide learners with the detailed feedback they need to improve their English writing skills, nor does it take into account that different ELLs can be at the same grade level, but in varying stages of learning English. (How can ELL writing be assessed, para 2)

Thus, there are limitations to using general criteria rubrics for writing assessment from the standpoints of both specificity and learner feedback. Particular problems are encountered while trying to develop descriptions in general criteria rubrics that accurately describe desirable writing qualities that divide writers into meaningful groups. For example, vague descriptive terms such as “well-organized,” “intermediate vocabulary,” and “frequent grammatical errors” are inevitable. Additionally, while they may be great time savers for writing instructors to assess writing skill with a reasonable degree of accuracy and within a narrow range of general criteria, these rubrics are essentially useless to ESL students from the point of view of feedback and possible improvement in their writing skills. As far as the students are concerned, general criteria rubric assessment is no more useful to them than letter grades in terms of informing them as to how they can improve their writing. Thus, if writing instructors intend to use rubrics both as scoring tools and for student feedback purposes, a more advanced rubric design is necessary.

Specific Criteria Rubrics

Specific criteria rubric (band assessment).

One option for creating a rubric to assess ESL writing proficiency is to use a specific criteria band assessment model (see Figure 2). The purpose of the band assessment model is to separate writing proficiency into three or more general classes, roughly equivalent to elementary, intermediate, and advanced groups. Establishing a range of numerical scores within each band further delineates individual variability. For example, an advanced score may range from between eight to ten, intermediate from four to seven, and elementary from one to three. This enables the rater to give appropriate credit for certain skill areas while at the same time recognizing weaknesses in others. As with the general criteria rubric, descriptions are provided that attempt to encapsulate desirable skills, the difference being that instead of an overall description of writing skills as seen with the general criteria model, the band assessment rubric contains separate criteria to be judged in turn, exclusive of other criteria. Thus, for example, using the Figure 2 rubric, a writer can be assessed as having high-order skills in Content and Organization, while at the same time only intermediate-level skills in Unity and Coherence and perhaps even elementary skills in Grammar and Vocabulary. This approach has two advantages: it allows the rater to focus on specific linguistic domains within the writing and, if used as an instructional rubric (Andrade, 2000), provides the student with information about specific areas within their writing upon which to concentrate.

The rubric shown in Figure 2 was developed to assess paragraph writing proficiency. To accomplish this objective, four criteria categories were established: Content and Organization; Unity and Coherence; Grammar and Vocabulary; and Format and Presentation. Due to the limited length and scope of the paragraph in academic writing, the criteria were established in pairs that I felt often work in unison. For example, Content and Organization often complement one another

to deliver a well-reasoned and well-supported message. Likewise, Unity and Coherence are both connected with comprehension and readability. Furthermore, Grammar and Vocabulary are complementary elements of writing with regard to linguistic accuracy.

Naturally, the number of criteria contained in a rubric and the relative scoring weight assigned to that criteria is a matter for the rubric creator to decide. Thus, more weight may be assigned to Content in content-based courses or courses utilizing the communicative approach. Alternatively, if the focus of the course is on grammatical accuracy, for instance, a greater scoring weight can then be assigned to the Grammar criteria. Making such decisions as to the criteria to be measured and the relative weight assigned to each can thus be done in accordance with course objectives.

Specific criteria rubric (paradigm assessment).

Writing can also be assessed by using a specific criteria paradigm assessment model (see Figure 3). The purpose of this model is to hold the writing up to a set of ideals, or paradigms, for individual linguistic categories similar to those used in the specific criteria band assessment model. With the paradigm assessment model, a number of specific, desirable writing features are grouped together under one category. For example, with the rubric presented in Figure 3 for assessing ESL paragraph-writing proficiency, features such as the inclusion of a properly written topic sentence, use of specific details for support, and use of thoughtful and interesting content are included under the Content & Organization category. Similarly, writing features such as correct spelling, correct grammar, and good word choice are listed under the Grammar & Vocabulary category. To help assess a piece of writing, then, this model lists the salient features of good writing that are being covered during a particular course or which represent the generally agreed upon body of desirable elements for a particular unit of organization, such as a paragraph, a five-paragraph essay, or an academic research paper.

The advantage to using a paradigm assessment model is that it utilizes the brevity and efficiency of the band assessment model while at the same time allowing for a higher degree of specificity in the descriptions listed under the various criteria. The paradigm model also does away with the arbitrary gradations of quality with regard to student writing assessment seen in the band assessment model. In other words, the students' work is held up for comparison with what constitutes an ideal version of the type of writing to which the student is expected to conform – a sort of checklist of pertinent features – thereby reducing the number of vague or imprecise descriptions such as “Near Native,” “Advanced,” “Upper Intermediate,” “Elementary,” and so on that one is occasioned to write when attempting to create graded bands of writing proficiency. Quite simply, a numerical value, for example, from a scale of one to ten, is determined for each of the criteria listed based upon the student's relative strength in that area. Holding the student up to an independent standard for each skill area described in the criteria helps to reduce the possibility that

an unfavorable assessment under one set of criteria would adversely affect accurate assessment in another. Thus, a poor performance under the Content & Organization criteria, for example, would not necessarily prejudice the educator toward assessing poor performance in another domain, as long as the standards in those other criteria are adhered to independently of one another. It may also work to counter preconceived notions on the part of the instructor as to the quality of a student's work based on recollections of previous work submitted.

Another advantage of the paradigm rubric is with respect to student feedback. In this regard, the higher degree of specificity used in the rubric descriptions makes it easier for students to use the rubric assessment to improve their writing skills. As Andrade (2005) notes, instructors will still want to provide written and verbal feedback in addition to rubric assessment. However, she goes on to point out that:

...if [instructors] were to simply circle boxes on a rubric and give it back with an assignment, [they] would still be providing more feedback about strengths and weaknesses of the work than had [they] just assigned a letter grade, and it would not take [them] any longer. (p. 29)

Accordingly, providing students with a more specific assessment of their writing skills based on descriptive paradigms has the additional benefit from the teacher's perspective of substantial time savings and efficiency. This last benefit accrues to both the instructor and the student in that it reduces the fatigue and ennui that could plague an instructor faced with assessing a large number of individual pieces of writing.

Conclusion

I have argued that rubrics can be a valuable tool for assessing student writing in several respects. First, they allow instructors to coordinate their assessment method with their course's original goals and objectives. Moreover, by creating a set of assessment guidelines rubrics establish consistency in writing assessment, both from the point of view of the individual instructor and from that of instructors using a unified curriculum or a common set of writing objectives. This reduces the amount of subjective decisions as to quality and content of a piece of writing to which an instructor may be prone. Finally, more specific rubrics can be used to provide useful and constructive feedback to students, thereby enabling them to focus on those areas of their writing that are in need of improvement.

Research into the use of scoring rubrics has found that the reliability of performance assessment can be improved through the use of scoring rubrics, and furthermore, that rubrics can both promote learning and improve the quality of instruction (Jonsson & Svingby, 2007). Rubrics provide the potential for enhancing objectivity and consistency in writing assessment if designed properly and in accordance with an instructor's course goals and objectives. Accordingly, their use in writing

assessment can improve the educational experience from the standpoint of both the instructor and the student.

Figure 1 – General Criteria Assessment Model (Paragraph Rubric)

5	Very well-organized writing with a logical order of presentation. The main idea is clearly expressed in a topic sentence and supported by sentences using specific examples, explanations, reasons, etc. Paragraph shows unity throughout with good use of transition words. A variety of sentence types are used, including simple, compound, and complex sentences. Higher-level vocabulary is used. Few, if any, grammatical or mechanical errors appear.
4	Essentially well-organized, coherent writing. The main idea is expressed in a topic sentence and developed with supporting sentences, although support may be vague and some detail lacking. Paragraph generally shows unity but some irrelevancies and segues may occur. Attempts to use a variety of sentences although not always successfully. Low intermediate vocabulary is used. Some grammatical or mechanical errors appear.
3	Writing shows weaknesses in organization and may lack coherence. The main idea may not be clearly expressed in a topic sentence or else is not adequately supported. Paragraph shows a lack of unity and contains several irrelevancies. Inadequate use of transition words. Simple sentences used with occasional attempts at sentence variety. Some intermediate-level vocabulary is used. Frequent grammatical or mechanical errors appear.
2	Writing very poorly organized and lacking in coherence. The main idea is not clearly expressed in a topic sentence and is not adequately developed. Little to no paragraph unity is evident and numerous irrelevancies occur. Transition words are infrequent or else incorrectly used. Simple sentences and fragments predominate. Mostly high-frequency vocabulary is used. Numerous grammatical or mechanical errors that interfere with understanding.
1	Writing not in basic paragraph form. Shows little to no attempt at organization. Consists mainly of a random series of incoherent or irrelevant sentences that do not support a main idea. Sentences may be repetitive or so vague as to lack any real meaning. Rudimentary vocabulary primarily used. Serious spelling and/or grammatical errors and lack of transition words make it very difficult to understand any intended message.

Figure 2 – Specific Criteria Band Assessment Model (Paragraph Rubric)

	Content & Organization	Unity & Coherence	Grammar & Vocabulary	Format & Presentation
10	Well-developed writing with main idea clearly stated in a topic sentence and supported with specific details including examples, explanations, reasons, etc. Content is informative and/or persuasive.	10 Arranged in some logical order (time order, spatial order, or order of importance).	10 Correct spelling and punctuation. Good use of higher-level grammatical forms and a variety of sentence patterns, including compound and complex sentences.	5 Margins, indentation, and spacing are all correct; paper is in correct paragraph form. Inclusion of all particulars (name, student number, class number, and title). 4 Very neat and clean presentation.
9	Moderately well-developed writing that generally revolves around a main idea. Reasonable support is provided, but may not always be specific. Adequate content that conveys meaningful ideas.	9 Supporting sentences are all directly connected to the main idea. Good use of transition words to link ideas together.	9 Good use of higher-level, descriptive vocabulary.	3 Writing generally conforms to formatting guidelines although not perfectly. Most particulars are included but some may be missing or incorrectly placed. Overall acceptable presentation.
8		8 Arranged around a reasonably discernible central idea.	8 Spelling and punctuation are generally correct, but areas of weakness may be evident. Acceptable grammar with some errors but generally comprehensible. Vocabulary is appropriate to task.	
7		7 Supporting sentences are provided that are generally clear and relevant. Some attempt at using transition words to link ideas together.	7 Spelling and punctuation are generally correct, but areas of weakness may be evident. Acceptable grammar with some errors but generally comprehensible. Vocabulary is appropriate to task.	
6	Poorly-developed writing with little attempt at paragraph organization. Support is often superfluous, repetitious, or irrelevant. Content is careless, uninteresting, or hurried.	6 Reasonable support is provided, but may not always be specific. Adequate content that conveys meaningful ideas.	6 Careless spelling and punctuation errors. Over-reliance on simple grammar and sentence patterns which are often incorrect in execution. Simple, everyday vocabulary that lacks precision and variety.	1 Writing disregards many paragraph formatting guidelines. Particulars are either missing or incorrectly placed. Sloppy or careless presentation. May include handwriting, tears, etc.
5		5 Writing is largely incoherent and rambling. Support, where evident, generally makes use of random, unconnected sentences. Little to no attempt to link ideas semantically.	5 Careless spelling and punctuation errors. Over-reliance on simple grammar and sentence patterns which are often incorrect in execution. Simple, everyday vocabulary that lacks precision and variety.	
4		4 Supporting sentences are provided that are generally clear and relevant. Some attempt at using transition words to link ideas together.	4 Spelling and punctuation are generally correct, but areas of weakness may be evident. Acceptable grammar with some errors but generally comprehensible. Vocabulary is appropriate to task.	
3	Poorly-developed writing with little attempt at paragraph organization. Support is often superfluous, repetitious, or irrelevant. Content is careless, uninteresting, or hurried.	3 Writing is largely incoherent and rambling. Support, where evident, generally makes use of random, unconnected sentences. Little to no attempt to link ideas semantically.	3 Careless spelling and punctuation errors. Over-reliance on simple grammar and sentence patterns which are often incorrect in execution. Simple, everyday vocabulary that lacks precision and variety.	1 Writing disregards many paragraph formatting guidelines. Particulars are either missing or incorrectly placed. Sloppy or careless presentation. May include handwriting, tears, etc.
2		2 Supporting sentences are provided that are generally clear and relevant. Some attempt at using transition words to link ideas together.	2 Spelling and punctuation are generally correct, but areas of weakness may be evident. Acceptable grammar with some errors but generally comprehensible. Vocabulary is appropriate to task.	
1		1 Arranged around a reasonably discernible central idea.	1 Spelling and punctuation are generally correct, but areas of weakness may be evident. Acceptable grammar with some errors but generally comprehensible. Vocabulary is appropriate to task.	
	Content & Organization ____ / 10	Unity & Coherence ____ / 10	Grammar & Vocabulary ____ / 10	Format & Presentation ____ / 5
Student Name: _____ Student Number: _____			TOTAL SCORE _____ / 35	

Figure 3 – Specific Criteria Paradigm Assessment Model (Paragraph Rubric)

	Content & Organization	Unity & Coherence	Grammar & Vocabulary	Format & Presentation
10	Topic Sentence - States topic and controlling idea - Neither too general nor too specific	10	10	
9		Arranged in a logical order (e.g., time order, spatial order, order of importance)	Correct spelling Correct punctuation	Proper use of margins
8	Supporting Sentences - advance the main idea through sufficient use of supporting points	8	Correct capitalization	5
7	- provide specific detail (e.g., examples, explanations, reasons, statistics, locations, descriptions)	Paragraph expresses a consistent point of view/train of thought	Correct grammar - verb tenses - singular vs. plural forms	Proper paragraph indentation
6		Use of transition or linking words (e.g., first, for example, however, furthermore, in other words)	- subject verb agreements - use of function words (i.e., conjunctions, prepositions, articles)	Double spaced
5	Concluding Sentence - restates the main idea or adds a final comment	5	5	4
4		All supporting sentences refer to the main idea (i.e., no irrelevant sentences)	No run-on sentences, fused sentences, or comma splices	Correct font style and font size
3	Thoughtful, interesting content appropriate to assigned task	3	High level vocabulary appropriate to content.	Inclusion of title in correct position
2		Proper use of pronouns, synonyms, and references as applies to referents	Good, varied word choice	Inclusion of name, class, and student number
1	Achieves paragraph objective to persuade, inform, or entertain	1	Good use of modifiers (i.e., adjectives and adverbs) and/or relative clauses.	Neat, clean presentation
		Correct use of coordinating and subordinating conjunctions		Fulfills minimum word or sentence requirement
	Content & Organization ____ / 10	Unity & Coherence ____ / 10	Grammar & Vocabulary ____ / 10	Format & Presentation ____ / 5
Student Name: _____ Student Number: _____			TOTAL SCORE ____ / 35	

References

- Andrade, H.G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5): 13-18
- Andrade, H.G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1): 27-30.
- Brown University (2015). Teaching diverse learners: How can ELL writing be assessed in the classroom? Brown University Education Alliance. Retrieved on August 20, 2015 from <http://www.brown.edu/academics/education-alliance/teaching-diverse-learners/question-v>
- Carnegie Mellon (n.d.). Rubrics – Teaching excellence & educational innovation. Retrieved on August 20, 2015 from <https://www.cmu.edu/teaching/design/teach/rubrics.html>
- Jonsson, A., & Svingby, G. (2007). The use of rubrics: Reliability, validity and educational consequences. *Educational Research Review* 2: 130-144.
- Rogers, B. (2007). *The Complete Guide to the TOEFL® Test: iBT Edition*. Boston: Heinle Cengage Learning.

(2016.9.23 受稿, 2016.11.25 受理)