Rubrics: How To and How Not To
by Marcella A. Farina and Nicole Hammond Carrasquel

Assessing language skills is a vital part of English language teaching and is equally important for language learners. The assessment process often includes the use of rubrics in order to obtain a more objective classification of learner ability, in both strengths and areas for improvement, and to help guide students toward mastering course objectives. While rubrics have been commonly used among language educators for a fairly long time, understanding, creating, and applying rubrics effectively continues to be challenging for not only novice but also experienced teachers.

The Basics of Rubrics

In general terms, we usually think of rubrics as being either holistic or analytic, each with its own advantages and disadvantages. Holistic rubrics usually consist of Likert-type ratings with broad descriptors that result in a single score representing overall language competency. They are easy to use in that they examine language proficiency in broad terms and are efficient in particular for large scale assessment, such as placement testing. However, this macro-type evaluation leaves educators with little or no knowledge about individual student proficiency, nor does it provide critical feedback to inform teaching and learning objectives (Hamp-Lyons, 1991). Other drawbacks of holistic rubrics relate to inter-rater reliability, where multiple raters interpret and apply the rubric criteria in different ways, and intra-rater reliability, where one rater's perception and application of the rubric criteria changes over time.

Analytic rubrics, on the other hand, consist of individually scored subcategories identifying more specific criteria. Although analytic rubrics also generate a composite score of proficiency, they also provide valuable evaluation feedback on specific language areas. This added information about the learner more directly guides curricular decisions and can better serve both educators and students throughout the learning process. Of course, analytic scoring is much more time-intensive, but research shows that it is typically more consistently applied (Hamp-Lyons, 1991).

Additional Considerations

Other factors can also impact the effectiveness of rubric use with language assessment, such as length, grammar complexity, word choice, error types, and text formats. One example of these factors comes from a study by Powers, Fowles, Farnum, and Ramsey (1994), which reported a statistically significant difference in holistic scoring for handwritten over word-processed texts. Their results showed that, compared to word-processed texts, handwritten language samples appeared longer on the page, benefited from obscure penmanship to mask errors, and revealed learner self-correction marks. These findings demonstrate that several and varied factors impact the effective use of rubrics for language performance assessment.

Preparing the Rubric

For teachers, the first daunting task of implementing rubrics is the creation of the rubric itself, and there are several questions that need to be addressed beforehand:

- What is the task/function being assessed?
- How many areas should be included in the rubric?
- What is the appropriate point-range?

Aspects that can render rubrics ineffective are descriptors that are either too vague or overlap conceptually and point ranges that contain too many or too few subvalues. Also key to developing an effective rubric is that the language level of the rubric be appropriate to the audience. This is particularly true for rubrics intended for use with beginning-level English learners. In other words, the language used within the rubric should be geared toward the English proficiency level of the students with whom it is being used so that they can easily understand evaluation criteria and appropriately interpret expectations (Griffith & Lim, 2012, p. 5).

Sample Rubric: Answering *Wh-* Questions

Below (Figure 1) is an example of an analytic rubric designed for beginning level, adult English learners enrolled in a grammar course. The activity asks the learner to write one complete sentence to answer each of the five *Wh-* questions presented in the task. In applying the rubric, each sentence is assessed based on the following criteria and corresponding values. Note that it is also possible for rubric criteria to reflect varying point ranges.

The student…	2	1	0
…writes a complete sentence (1 subject, 1 verb).	has both	has one	has neither
…uses the correct verb tense.	0 errors	1 error	2+ errors
…gives information that answers the *Wh-* question.		yes	no

Figure 1. Sample rubric

Using the Rubric

Once we have created a rubric we believe to be valid for the language skills we are assessing, the next challenge is implementing the rubric in the curriculum, which leads to another essential aspect to consider. How do I prepare my students for the inclusion of the rubric in the learning?

Ideally, teachers should discuss with the class rubric criteria and the rubric point range before they apply the rubric to the actual assessment of student work. This can easily be done by incorporating into the lesson a peer activity whereby students use the rubric collaboratively and learn how to apply the rubric criteria in the form of self- and peer-evaluations. Teachers should also be vigilant that rubrics are designed to address only areas that have been explicitly taught and practiced in the classroom. It is unrealistic and ineffective to apply a rubric pointed on assessing, for instance, article usage if articles have not been addressed, or the correct pronunciation of *all* English sounds if only certain phonemes have been taught. These considerations help transform rubrics into transparent tools that enrich the language learning experience.

Strengthening the Rubric

Calibrate With Colleagues

To further strengthen the effectiveness of rubric use in assessment, it is also important for teachers to think in collaborative terms. If the rubric were given to another teacher to use, would it be perceived and applied similarly? How do rubric users distinguish labels, such as "somewhat accurate" from "generally accurate"?

Even rubrics used in standardized testing, such as TOEFL iBT and IELTS, include extensive protocol to ground rater perceptions and minimize differences in applying the criteria. One way for us to reduce these variations is to arrange with a colleague to "partner score" (trade scoring duties for your respective classes but use the same rubric). Another strategy is "duo scoring," where each sample is scored by two teachers and the scores are compared for similarity. For optimal results with these scoring practices, language samples should be anonymous, and scoring should be independent.

Create an Exemplar Bank

Another technique for bringing rubric perceptions closer together is to establish over time a bank of exemplars. To do this, teachers gather samples of student writing that characterize the various criteria referenced on the rubric and use the samples to ground perceptions about the rubric criteria and features of the point-range.

Although these strategies may take a little time to implement and standardize in the beginning, the long-term gains are greater efficiency and cohesiveness of colleagues applying rubric criteria within a school or program.

Closing Thoughts

All in all, in the everyday classroom, teachers are confronted with the difficult task of creating valid and reliable ways to efficiently evaluate language learning, and rubrics can facilitate this process greatly if several factors are thoroughly considered before, during, and even after the rubric implementation process. According to a meta-analysis by Jonsson and Svingby (2007), reliable scoring is improved if performance assessment rubrics are utilized, and is strengthened further if those rubrics are analytic, topic-specific rubrics and include benchmarks and rater training procedures. Their research suggests that rubric use bolsters clarity in assessment expectations and enhances learning and instruction through enriched feedback. Being aware of these considerations from the start will help educators devise rubrics to better evaluate outcomes and more wholly support teaching and learning.

References

Griffith, W. I., & Lim, H. Y. (2012). Performance-based assessment: Rubrics, web 2.0 tools and language competencies. *MEXTESOL Journal*, *36*(1), 1–12.

Hamp-Lyons, L. (1991). Scoring procedures for ESL contexts. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 241–276). Norwood, NJ: Ablex.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130–144.

Powers, D. E., Fowles, M. E., Farnum, M., & Ramsey, P. (1994). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement, 31*(3), 220–233.

_____

[Dr. Marcella Farina](), assistant professor in TESOL at the University of Central Florida, has more than 30 years teaching experience in varied adult ESL/EFL settings, with an extensive background in preacademic English program administration and instruction. Her teacher education background ranges from elementary school to adult education internationally. Her research interests focus on pronunciation teaching and learning and developing synchronous online language teaching models and strategies.

[Nicole Hammond Carrasquel]() has been teaching ESL to adults since 2001. She has her MA in TESOL and is currently working toward her PhD in TESOL. She taught for 11 years at an intensive English program at the University of Central Florida where she was a full-time instructor and member of the curriculum committee. She has developed curricula for several short-term programs and international teacher trainings. Her research interests are in foreign accents and pronunciation.