Thyme room
1.10-1.35pm

Applying a rubric development cycle for assessment in higher education: an evidence-based case study of a science communication module

Brenda YUEN* and Sirinut SAWATDEENARUNAT
Centre for English Language Communication, National University of Singapore
*elcbypl@nus.edu.sg

Although empirical studies on the use of rubrics have been undertaken in a wide range of disciplines and for various purposes in higher education, surprisingly little attention has been paid to the process of rubric development, especially in the assessment of science communication, to establish the quality of rubrics, i.e. *validity* (the extent to which scores reflect the underlying variables of interest) and *reliability* (consistency of scores across repeated measurements). This paper presents a case study on how a rubric development cycle from the design and application to validation was incorporated into the development of an analytic rubric for a science communication module for NUS science undergraduates.

To create a rubric that suits the needs and requirements of this educational context, we implemented a rubric development cycle (Yuen & Sawadeenarunat, 2018), which consists of three key stages: 1) a four-step process to rubric construction; 2) rubric application with tutors and students; and 3) a mixed-method rubric validation. These three stages are in a loop that can be iterated to enhance rubric quality for assessment and grading, providing effective feedback, and promoting student learning.

In the process of rubric construction, Popham's (1997) three key features of a rubric, namely *evaluative criteria, quality definitions* and *scoring strategies*, and Dawson's (2017) identification of 14 rubric design elements were considered. Based on a four-step process to rubric construction, we developed an analytic rubric for a science news article task which measures students' writing ability to communicate scientific discoveries to non-specialist readers. The rubric adopts an 8-point rating scale (i.e. *scoring strategies*) and has four *evaluative criteria* of the same weighing: 1) *accessibility/readability*; 2) *organisation of ideas*; 3) *significance of the key finding*; and 4) *language strategies to engage readers.*

In terms of rubric application, both tutors and students received face-to-face training to understand the *quality definitions,* and the standards of performance levels. Nine trained tutors were involved in the rating process in which they applied the rubric to assess student work and provide feedback. Students were also trained to use the rubric to assess the quality of science news articles. 334 science news articles were marked, 63 of which were double-marked.

The validation process adopted a mixed methods research design in which 4 types of data were collected and analysed: 1) tutors' ratings; 2) tutors' annotations of scripts; 3) tutors' semi-structured interview data; and 4) students' survey responses. To examine the psychometric quality of the rubric, ratings were analysed using the Many-facet Rasch Model (MFRM). The MFRM refers to a class of measurement models that extends the basic Rasch model (Rasch, 1980) by incorporating more variables or facets than the two that are typically included in a testing situation, namely students and items (Eckes, 2009). It is used to assess the degree of rater severity and consistency, correct scores for rater severity differences, and to examine the difficulty of items and functioning of the rating scale. The results of the MFRM analysis indicates that the rubric appears to be functioning well as the raters, items and rating scale function as intended by the model, but it needs revision by reducing rating scale categories. The qualitative analysis of tutors' feedback also supports revision and recommends refinement of the descriptors. The results from students' surveys show that students found the rubric useful in helping them understand their performance levels and in enhancing their writing performance. This case study that adopted the rubric development cycle not only provides direction for future revision of the rubric, but it also suggests implications for rubric development and validation in higher education.

Keywords

assessment; rubric design and validation; reliability and validity; Many-facet Rasch model

References

Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education, 42*(3), 345-360. http://dx.doi.org/10.1080/02602938.2015.1111294

Eckes, T. (2009). Many-facet Rasch measurement. In Takala, S. (Ed.), *Reference supplement to the manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (Section H)*. Strasbourg, France: Council of Europe/Language Policy Division.

Popham, W. J. (1997). What's wrong—and what's right—with rubrics. *Educational Leadership, 55*(2), 72-75. http://dx.doi.org/10.4135/9781446263549.n28

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Yuen, B. & Sawadeenarunat, S. (2018). *Assessing science communication competence: a rubric designed for popular science news writing*. Paper presented at Asia TEFL 2018 International Conference, University of Macau, Macau, China, 27-29 June 2018.