A RUBRIC FOR SCORING POSTSECONDARY ACADEMIC SKILLS

Marielle Simon & *Renée Forgette-Giroux,*
Faculty of Education, University of Ottawa

Today's assessment of postsecondary academic skills must take into account their comprehensive nature and their multiple facets (Biggs, 1995; Sadler, 1989). In this regard, the use of rubric is more likely to provide qualitative, meaningful, and stable appraisals than are traditional scoring methods. The stability of assessment results, however, rests on the scale's ability to lead to a common and uniform interpretation of student performance. The assessment of postsecondary academic skills on the basis of such a scale offers several advantages. First, it presents a continuum of performance levels, defined in terms of selected criteria, towards to full attainment or development of the targeted skills. Second, it provides qualitative information regarding the observed performance in relation to a desired one. Third, its application, at regular intervals, tracks the student's progress of his or her skill mastery. Finally, the choice of rather broad universal criteria extends the application to several contexts.

Despite its merits, however, the use of a generic descriptive scale at the postsecondary level is relatively recent and some difficulties need to be addressed. This paper has three objectives:

   a. to present the nature of a generic rubric used to assess postsecondary academic skills,
   b. to describe a preliminary application in a university setting, and
   c. to discuss observed related issues from a research point of view.

Nature of the rubric

The rubric for scoring academic skills is essentially qualitative and descriptive in nature and relies on criterion-referenced perspectives. It serves to appraise academic competencies such as the ability to critique, to produce scholarly work, to synthesize, and to apply newly acquired principles and concepts. It requires the use of criteria that best describe actual student products in a postsecondary setting. The criteria form the left-hand column of the two-way table format and the horizontal continuum contains headings indicating four increasing levels of performance towards competency mastery (Wiggins, 1998).

The use of the scale involves the acts of scoring, interpreting, and judging. (Forgette-Giroux, & Simon, 1998; Simon, & Forgette-Giroux, 2000). Scoring occurs when one identifies, within the scale, and for each criterion, the cell description that most closely matches the observed performance. The interpretation consists of locating the column that best describes the level of skill mastery. Judging means comparing the identified or observed performance level to a predetermined standard level.

Context of Application

The rubric discussed in this paper has evolved over the past five years but the latest, most generic version, was used within four graduate and two undergraduate level courses. Course enrollment varied from three to 30 students for a total of approximately 100 students. The courses were taught by the two authors, both experts in measurement and evaluation in education, and their topics related to research methodology or assessment. Given their theoretical nature, all courses were organized to assist students in their development and mastery of a single, carefully formulated academic skill, such as the ability to critically analyze a variety of research studies in education, to write a research proposal or report in education, and to assess student learning using current assessment methods and principles. Students were asked to assemble a portfolio that included scholarly works such as critiques, proposals, essays, manuscripts (Forgette-Giroux, et al., 1998). Practical assignments such as lesson plans, tests, performance assessments, were always accompanied by a structured critique. Students used the scale to self-assessed their portfolio for formative and summative purposes.

In this specific context, the performance levels, or anchors, are labeled as *good, very good, excellent*, and *exceptional*, to conform to the university approved grading scale. The five criteria are: Relevance, scope, accuracy, coherence, and depth. These criteria are commonly applied to scholarly writing by most manuscript review processes (NCME, PARE)[1], as are other attributes such as clarity, rigor, appeal, and strength of argument. The five criteria are also those found in the curriculum scoring rubrics mandated by the regional educational jurisdiction. The latter were to be learned by the

training teachers at the undergraduate level and eventually used in their future teaching environment.

During the repeated application of the scale to the various university level courses, three concerns arose that have also been noticed elsewhere, and which continue to interest researchers. The following sections describe these difficulties and present their tentative treatment in this particular university setting.

Scale levels (anchors) identification

When the stages of development or mastery of the targeted skills are not empirically grounded, the initial identification of the scale levels is often arbitrarily determined. Also, when courses are given for the first time, the lack of student work samples further complicates the scale level identification process. Some researchers define scale levels and criteria in a *post hoc* fashion, such as was the case with the National Assessment of Educational Progress (Burstein, Koretz, Linn, Sugrue, Novak, Baker, & Lewis Harris, 1995/1996). The difficulty with this approach is that it is context specific and students cannot be made aware of these parameters prior to the assessment. An alternative procedure is to select work from the student at hand, that is typical of the upper levels of the scale or of the standard level. Wiggins (1998) suggests that, given clear parameters around the intended use of the rubric, those criteria that make the most sense are chosen with an understanding that they may be constantly adjusted based on exemplary performances. In the university context described here, the first version of the scale was developed around the expected student performance at the level of *excellence*. As the course progressed, performance exemplars of that level were identified, distributed among the students, and used to refine the scale.

Specificity of descriptors

For the scale to be generic enough to be applied in a variety of university courses, the descriptors need to refer to a spread of performances at each level. On the other hand, there is a risk that these statements may be too general and thus lead to inconsistent interpretation of the data. In the study reported here, the descriptors were formulated based on criteria associated with the development of valued academic skills that are relatively independent of the course contents. These skills tend to combine declarative and procedural knowledge with scholarly writing. The universality and pertinence of the selected criteria in terms of academic and practical perspectives extended the applicability of the descriptors to a variety of courses at both undergraduate and graduate levels and ensured student endorsement. In addition, the formative assessment at least once during the course, allowed the students and their professor to mediate scale interpretation in order to produce stable results. Despite its early stages of development, the scale yielded average percent agreements of 75 % between professor and student ratings.

Qualitative rating versus quantitative scoring

Student bodies and administrative pressures stress the attribution of a letter grade or a quantitative score to ratings obtained using the descriptive scale. In assigning a score, the rubric loses its ability to provide detailed and meaningful information about the quality of the performance as reflective of a specific level of skill mastery. Within the study context, the university administration required the presentation of a letter grade. Its scale equates *Exceptional* with the letter A+, *Excellent* with A/A-, *Very good* with B+/B and *Good* with C+. Throughout each course, assessment results were communicated to the students primarily using descriptive statements based on the rubric, but a final letter grade had to be assigned at the end of the course for official transcript purposes. It is interesting to note that, in adopting the scale for their own courses, colleagues typically experienced the need to quantify their assessment using complex algorithms, medians, modes, averages. In doing so, they easily lost track of the object of assessment. It would appear therefore, that the transition toward a purely qualitative approach within certain administrative constraints, takes repeated applications, discussion, and much self-reflection.

Discussion and Conclusion

The rubric was initially conceived as a substitute for the numerical scale that became obsolete and unstable in its traditional application, particularly when assessing complex skills through performance assessments. Its usefulness in higher education, therefore, largely depends on its ability to lead to meaningful and stable assessment results. Relevancy and consistency of results refer to validity and reliability issues. Among some of the design considerations put forward by Arter (1993) in the selection of good criteria when constructing rubrics for performance assessments, the most relevant to postsecondary contexts are (a) the need for universal attributes, (b) the means for assessing both holistically and analytically, and (c) the identification of the main components of the object of assessment. Moskal and Laydens (2000) have proposed practical ways to address these issues. They equate evidence related to content with the extent to which the rubric relates to the subject domain, and construct-related evidence to the conceptualization of a complex internal skill. Criterion-related evidence, meanwhile, serves to indicate how well the scoring criteria match those found in practice. Given this rubric's generic nature and the focus on the assessment of academic skills, primary attention must be given to the production of construct-related evidence. This was achieved by linking the scale's criteria, anchors, and descriptors to the nature of the skill addressed by the rubric and expressed in terms of a single learning objective.

Interrater and intrarater aspects of reliability were greatly improved by attaching the rubric to the course outline and by clarifying its various components and use early in the course, by enabling the students to access high quality

exemplars, by providing regular qualitative feedback, by inviting the students to take part in mediation during formative assessments, and by requesting them to justify, in writing, their self-assessment based on specific references to their portfolio. It was important that this written rationale clearly support their perceived level of achievement. Written support of scoring decisions by the professor was also expected.

Given the exploratory nature of the study, many questions arise. However, five are of particular interest from both practical and research perspectives:

1. Would a scale based on a combination of the *post hoc* approach, of theoretical foundations of academic skills, and of samples of student work lead to increased validity?
2. Would a scale based on a combination of both qualitative and quantitative components lead to even greater consistency of results when assessing academic skills?
3. Is there an optimal class size to which this rubric can be applied most efficiently?
4. What type of teaching style is more likely to fit with the effective use of the scale?
5. Would the scale be as useful in courses focusing on content rather than skill development?

Research and dialogue on the obstacles and advantages of this approach are definitely needed to achieve some balance and to assist professional educators in addressing these issues when using the rubric within their own courses. Another dimension in need of further investigation would be to obtain evidence of convergent and discriminant validity. Finally, a rigorous, larger scale validation study of the universality of the criteria is also warranted if the scale is to become a widespread, valuable and valued tool in the assessment of postsecondary academic skills.

Footnotes

1 See http://ncme.ed.uiuc.edu/pubs/jem_policy.ace and http://pareonline.net/Review.htm

References

Arter, J. (1993). *Designing scoring rubrics for performance assessments: The heart of the matter*. Portland, OR.: Northwest Regional Education Laboratory. (ERIC Document Reproduction Service No. ED 358 143)

Biggs, J. (1995). Assessing for learning: Some dimensions underlying new approaches to educational assessment. *The Alberta Journal of Educational Research, XLI*(1), 1-17.

Burstein, L., Koretz, D., Linn, R., Sugrue, B., Novak, J., Baker, E.L., & Lewis Harris, E. (1995/1996). Describing performance standards: Validity of the 1992 National Assessment of Educational Progress achievement level descriptors as characterizations of mathematics performance. *Educational Assessment, 3*(1), 9-51.

Forgette-Giroux, R., & Simon, M. (1998). L'application du dossier d'apprentissage au niveau universitaire. *Mesure et évaluation en éducation*, 20(3), 85-103.

Moskal, B.M., & Leydens, J.A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation, 7*(10). Available online: http://pareonline.net/getvn.asp?v=7&n=10

Sadler, R. (1989). Specifying and promulgating achievement standards. *Oxford Review of Education, 13*(2), 191-209.

Simon, M., & Forgette-Giroux, R. (2000). Impact of a content selection framework on portfolio assessment at the classroom level. *Assessment in Education: Principles, Policy and Practice, 17*(1), 103-121.

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass Publishers.

APPENDIX

Descriptive scale: EDU5499 Current methods of student assessment in teaching and learning (graduate level course).

Learning objective: To be able to critically analyze the technical qualities of own assessment approaches.

Name: _____   Date:_____

Assessment   ☐ formative   Assessor: ☐ self

☐ summative   ☐ professor

CRITERIA	PERFORMANCE LEVELS (ANCHORS)			
	GOOD	VERY GOOD	EXCELLENT	EXCEPTIONAL
RELEVANCE Portfolio components address directly the learning objective.	Portfolio components are not necessarily related to the learning objective.	Portfolio components are somehow linked to the learning objective.	Portfolio components are directly linked to the learning objective.	The portfolio reflects exemplary high relevance to the learning objective

SCOPE All aspects of the learning objective and recommended readings are covered within the portfolio.	The portfolio partially reflects the various components of the learning objective.	Most elements of the learning objective and recommended readings are covered within the portfolio.	All aspects of the learning objective and all recommended readings have been dealt with within the portfolio.	The portfolio incorporates treatment of elements beyond the scope of the class and recommended references.
ACCURACY Current concepts, terms, principles, and conventions are used correctly and with clarity throughout the portfolio.	Learned concepts, terms, principles, and conventions are more or less used correctly throughout the portfolio.	The portfolio shows precision in the use of current concepts, terms and principles but relevant conventions not always followed.	The portfolio reflects correct and clear use of terms, concepts, principles, and conventions.	The portfolio demonstrates clear, correct, precise, and concise use of terms, concepts, principles and conventions.
COHERENCE Elements within and across the portfolio are logically and structurally linked together. Ideas are interconnected and are presented in a consistent fashion throughout the portfolio.	Elements and ideas are presented in a disconnected, rather piecemeal fashion.	Elements are somehow linked together but reflect some inconsistency across the portfolio.	Evidence of structural and internal consistency within and to some extent, across the portfolio.	The portfolio is highly and tightly organized. Ideas, concepts and principles are presented in a consistent fashion across the portfolio.
DEPTH The portfolio reflects a personal position supported by a rich analysis of relevant and high quality references.	The portfolio presents a position highly dependent on a superficial analysis of references.	The portfolio presents a position supported by some analysis of relevant references.	The portfolio reflects a personal position based on a deep and thorough analysis of relevant references.	The portfolio presents a personal position based on an integration of relevant and high quality references.

Descriptors: *Rubrics; Scoring; Student Evaluation; Test Construction

Citation: Simon, Marielle & Renée Forgette-Giroux (2001). A rubric for scoring postsecondary academic skills. *Practical Assessment, Research & Evaluation*, 7(18). Available online: http://PAREonline.net/getvn.asp?v=7&n=18.