



Developing Valid and Reliable Rubrics for Writing Assessment: Research and Practice

Author

Dr Keith Comer, University of Canterbury

Introduction

Writing assessment has long been a particular challenge for academic staff. A broad range of theoretical approaches have resulted quite varied practices, from traditional attempts to have single academics read through piles of student manuscripts to systematic, programme-wide efforts involving grading teams and multiple readers. The diversity of practices results from diverse perceptions and beliefs, training and levels of experience, influences and motivations, resources and institutional policies.

Fundamentally, these endeavours are intended to provide writing assessments that are valid and reliable. A valid method should provide accurate assessment of student learning in connection with desired outcomes, and a reliable assessment should offer fair and consistent evaluation of learning in connection with task requirements and expectations.¹ There is and will continue to be ongoing tension between these two aims. For example, while a multiple-choice test can provide perfectly reliable grading, what valid assessment of a student's learning with respect to academic writing could it offer? This guide explores how academic staff can pursue the aims of valid and reliable writing assessment in the context common disciplinary, programme and institutional challenges.

In attempting to provide more reliable and valid assessments of student learning, a number of academic writing programmes have endeavoured to shift from individual practices conducted in isolation to small group or even programme-wide marking approaches. These have included (singularly or in various combinations):



- portfolio assessment projects (from single to multiple course or even comprehensive undergraduate writing portfolios, initially on paper and now view e-portfolios)
- high stakes 'final essays' marked by staff teams in ways parallel to college entrance examinations
- team-taught courses
- comprehensive writing across curriculum (WAC) or writing in the discipline 'studios' (now typically supported through learning managements systems [LMS] such as Blackboard, Moodle, textbook publisher online systems, or other alternatives)
- technology supported distributed grading. (ENDNOTE w/ references for these).

An integral part of these approaches generally involves the use of assessment rubrics (also known as 'marking schemes' or 'marking guides' in NZ, Australia and the UK).² Such uses of rubrics arise partly from historical practices involving large-scale, university entrance examinations – in which inter-rater reliability is essential – and partly from efforts designed to minimize marking time, thereby providing more timely feedback or the capacity to allow for multiple readers. Gradually the use of rubrics and related marking guides has spread to nearly every academic discipline and into a range of primary, secondary and tertiary contexts.³

Written communication VALUE rubric

Below is an example of a rubric for writing assessment from T. L. Rhodes' 2010 *Assessing Outcomes and Improving Achievement: Tips and Tools Using Rubrics*.⁴

For more information, please contact: value@aacu.org

Definition

Written communication is the development and expression of ideas in writing. Written communication involves learning to work in many genres and styles. It can involve working with many different writing technologies, and mixing texts, data, and images. Written communication abilities develop through iterative experiences across the curriculum.

Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.

Employing rubrics can be especially effective in assessment contexts to address these issues: creating cooperative approaches with teachers of widely disparate levels of experience, fostering shared learning outcomes that are evaluated consistently, providing timely feedback to students, and integrating technology-enhanced processes with such rubrics can provide for greater flexibility in assessment approaches, provide for multiple readers of student writing, and more fully distinguish 'teachers' from 'evaluators'. This guide examines each of these issues in turn, based upon experiences and data from a project in the United States in 2006-2008.

[Dimension]	[Scale or Standard of Work]			
	Capstone 4	Milestones 3 2		Benchmark 1
Context of and Purpose for Writing <i>Includes considerations of audience, purpose, and the circumstances surrounding the writing task(s).</i>	Demonstrates a thorough understanding of context, audience, and purpose that is responsive to the assigned task(s) and focuses all elements of the work.	Demonstrates adequate consideration of context, audience, and purpose and a clear focus on the assigned task(s) (e.g., the task aligns with audience, purpose, and context).	Demonstrates awareness of context, audience, purpose, and to the assigned tasks(s) (e.g., begins to show awareness of audience's perceptions and assumptions).	Demonstrates minimal attention to context, audience, purpose, and to the assigned tasks(s) (e.g., expectation of instructor or self as audience).
Content Development	Uses appropriate, relevant, and compelling content to illustrate mastery of the subject, conveying the writer's understanding, and shaping the whole work.	Uses appropriate, relevant, and compelling content to explore ideas within the context of the discipline and shape the whole work.	Uses appropriate and relevant content to develop and explore ideas through most of the work.	Uses appropriate and relevant content to develop simple ideas in some parts of the work.
Genre and Disciplinary Conventions <i>Formal and informal rules inherent in the expectations for writing in particular forms and/or academic fields (please see glossary).</i>	Demonstrates detailed attention to and successful execution of a wide range of conventions particular to a specific discipline and/or writing task (s) including organization, content, presentation, formatting, and stylistic choices	Demonstrates consistent use of important conventions particular to a specific discipline and/or writing task(s), including organization, content, presentation, and stylistic choices	Follows expectations appropriate to a specific discipline and/or writing task(s) for basic organization, content, and presentation	Attempts to use a consistent system for basic organization and presentation.
Sources and Evidence	Demonstrates skillful use of high-quality, credible, relevant sources to develop ideas that are appropriate for the discipline and genre of the writing	Demonstrates consistent use of credible, relevant sources to support ideas that are situated within the discipline and genre of the writing.	Demonstrates an attempt to use credible and/or relevant sources to support ideas that are appropriate for the discipline and genre of the writing.	Demonstrates an attempt to use sources to support ideas in the writing.
Control of Syntax and Mechanics	Uses graceful language that skillfully communicates meaning to readers with clarity and fluency, and is virtually error-free.	Uses straightforward language that generally conveys meaning to readers. The language in the portfolio has few errors.	Uses language that generally conveys meaning to readers with clarity, although writing may include some errors.	Uses language that sometimes impedes meaning because of errors in usage.

The issue of teaching experience

Many academic writing programmes, whether situated within English or other individual departments or involving multiple departments in writing across the curriculum (WAC) or writing in the disciplines (WID) arrangements, employ a range of staff in teaching and assessing student writing. These can routinely include academic staff with postgraduate qualifications on continuing or fixed-term contracts, fixed-term

tutors and postgraduate students (at masters or doctoral levels), and sometimes advanced undergraduates or even secondary teachers providing 'dual enrolment' or 'advanced study' courses for tertiary credits. Some staff may have years or decades of experience teaching writing; others may be early career academics; and still others may be postgraduates teaching – independently, in mentor programmes, or as tutors – for the first time.

The use of common assessment rubrics with teachers of markedly different levels of experience can help minimize disparate assessment practices.⁵ For example, an approach piloted at a university, a polytechnic and a secondary school 'dual enrolment' program in Idaho, USA included a total of 14 teachers from all groups noted above, with teaching experience ranging from a single prior course to over 20 years of university instruction. During two years they taught 21 sections (or 'streams') of two university-credit academic writing courses at three geographically separated locations, with over 430 students who wrote 800+ essays, and those essays required some 2,000+ assessments. Common rubrics – jointly developed by all teaching staff involved – helped focus attention on the instructional aims rather than the levels of instructor experience. A vital component of this process is to ensure that teachers providing the same course engage in on-going conversations about the assessment process as well as the evaluation instruments or rubrics involved.⁶

In addressing different levels of teaching experience, structured and semi-structured dialogues between teachers centred on the writing tasks and the assessment processes for those tasks are essential. While some of these conversations may be facilitated online, occasional face-to-face meetings are invaluable and at a minimum should be arranged before term, before and after the first major assessment, and following final marking for the term.

Fostering shared learning outcomes and consistent evaluation

Ideally, assessment rubrics should be developed by the teachers employing them. However, there may be constraints on this process. Programmes and institutions often have specific learning outcomes identified, and those outcomes need to form the basis for assessments. Additionally, accreditation agencies or internal/external programme reviews may have set requirements that must be addressed via assessment policy and practice. Yet whether learning outcomes can be developed or revised by teaching staff engaged in assessment, or if those staff must integrate pre-existing requirements and guidelines or templates in their assessments, the use of rubrics can provide opportunities for fostering shared practices.

The key here is that the learning outcomes for a course provide the basis for any assessment rubrics. Particularly important to note is that not every outcome need be assessed for every essay or assessment task. For example, it can be beneficial to focus on one or two learning outcomes at a time, particularly with respect to assessments early in a term.⁷ Time and space do need to be made available

for teachers to meet and collaborate on the rubrics that will be used to assess which learning outcomes are involved for specific assignments.

Establishing this shared agreement on learning outcome assessment begins to create the conditions for more consistent evaluation practices. In a parallel situation to the one noted above – but with academics from a range of disciplines – A. J. Levi and D. D. Stevens record how “two English professors, an anthropologist, a historian, a chemist, and even a high school teacher offering the class as an Advanced Placement option” worked to establish a common rubric for a task shared between their very different courses. The shared goals – evaluating student learning in connection with established learning outcomes – provide the foci for developing the rubrics.

In the Idaho project, the main rubric employed was designed in such a collaborative process: three months ahead of term, project teachers were provided with a number of sample rubrics. These were reviewed independently, and then teachers met to discuss and devise a shared model – which involved an adaptation of a rubric used for holistic portfolio assessment conducted as part of an accreditation review (and was, in its prior form, an adaptation of a rubric used in another university for similar purposes). The final form of whatever rubrics are developed is less important than the shared process whereby it is shaped and decided upon. The teachers involved are thereby more invested in clarifying and establishing direct connections between their collaboratively developed rubrics and course learning outcomes. Consequently, the validity of assessments produced tends to improve, since assessments are conducted from shared perceptions and understandings of the specific learning that is being measured, rather than individual interpretations of such outcomes that are carried out in isolation.

With shared agreement on the measurement of learning outcomes, reliable marking practices are more easily developed. Considerable research demonstrates the improved consistency in marking that rubrics can help foster. This reliability has frequently been an important factor in their adoption for diverse evaluations, from large-scale, standardized exams that grade single, high-stakes samples of students’ writing to individual portfolio marking involving multiple years and diverse collections of student work.

When beginning to work with shared rubrics, all teachers involved must engage in some preliminary ‘norming’ or ‘moderating’ sessions. These require working with the rubric and a number of samples of student writing in establishing the desired degree of inter-rater reliability – that is, how well do the various individuals taking part assign the same assessment marks? Generally there will need to be at least two such norming/moderation sessions. Although participants can employ the rubrics involved independently prior to such meetings, following each session there typically needs to be an open discussion of results and exchange of feedback involving the variation in scores that arises. Most groups of teachers can establish satisfactory levels of marking consistency in two or three such sessions, typically involving 10-12 student writing samples. Depending upon the length of the term or period during which rubrics will be employed collaboratively, it can also be helpful to work through a ‘mid-term’ norming/moderation session. This can serve to reinvigorate the assessment process and help maintain more consistent approaches as end-of-term marking approaches (and generally increases).

The need for timely feedback to students

Student learning benefits from timely and well-designed feedback. With generally increasing class sizes and competing pressures (teaching, research, service...) on time available for marking, implementing a rubric-based approach can help provide faster as well as valid and reliable feedback. This time-savings can become substantial when technology-enhanced options are pursued. For example, L. Anglin and K. Anglin's 2008 study demonstrated that computer-assisted grading rubrics could be completed from 200%-350% faster than more traditional approaches. This is supported by practices developed from the Idaho project, which through the use of online rubrics reduced assessment marking from an average of 20 minutes per student essay to less than 10 minutes per essay. Time saved by adapting technology-enhanced marking practices can then be used elsewhere – for example, in student-teacher conferences or tutorial support. It can also be used to dramatically transform the marking process.

As part of the Idaho project, the time-savings developed in this manner was redirected to allow for multiple reviews of student writing, in a process that paralleled peer review for academic journals. The distributed grading software application developed as part of the project allowed for student essays to be submitted online and graded by at least two reviewers other than the instructor of record for each student, with evaluations returned online immediately. If the first two evaluations disagreed by more than a pre-set standard (e.g., 10%), a third reviewer was engaged in the process. Averaged scores (either numerical, letter, or of the P/NP variety) from the two closest were returned to the student immediately upon completion, along with complete commentary from both reviewers.

Instead of “writing for the teacher,” students became responsible for meeting the challenging standards of audiences composed of multiple reviewers throughout the course of the term and made possible through the technology-enhanced distributed grading model. While the assignment of reviewers was initially done randomly, subsequent essays from the same student were then be evaluated by as many different reviewers as possible during the semester. For example, if a course group involved 12 reviewers, any individual student could offer a minimum of four essays before any single reviewer evaluates that student's work more than once.

Student learning benefits from timely and well-designed feedback. With generally increasing class sizes and competing pressures (teaching, research, service...) on time available for marking, implementing a rubric-based approach can help provide faster as well as valid and reliable feedback. This time-savings can become substantial when technology-enhanced options are pursued. For example, L. Anglin and K. Anglin's 2008 study demonstrated that computer-assisted grading rubrics could be completed from 200%-350% faster than more traditional approaches. This is supported by practices developed from the Idaho project, which through the use of online rubrics reduced assessment marking from an average of 20 minutes per student essay to less than 10 minutes per essay. Time saved by adapting technology-enhanced marking practices can then be used elsewhere – for example, in student-teacher conferences or tutorial support. It can also be used to dramatically transform the marking process.

As part of the Idaho project, the time-savings developed in this manner was redirected to allow for multiple reviews of student writing, in a process that paralleled peer review for academic journals. The distributed grading software application developed as part of the project allowed for student essays to be submitted online and graded by at least two reviewers other than the instructor of record for each student, with evaluations returned online immediately. If the first two evaluations disagreed by more than a pre-set standard (e.g., 10%), a third reviewer was engaged in the process. Averaged scores (either numerical, letter, or of the P/NP variety) from the two closest were returned to the student immediately upon completion, along with complete commentary from both reviewers.

Instead of “writing for the teacher,” students became responsible for meeting the challenging standards of audiences composed of multiple reviewers throughout the course of the term and made possible through the technology-enhanced distributed grading model. While the assignment of reviewers was initially done randomly, subsequent essays from the same student were then be evaluated by as many different reviewers as possible during the semester. For example, if a course group involved 12 reviewers, any individual student could offer a minimum of four essays before any single reviewer evaluates that student’s work more than once.

Considerations

This article does not attempt to settle the debate regarding validity, reliability and writing assessment. But it does explore how rubrics can be used to provide valid and reliable assessments with diverse groups of teaching staff. Further, when integrated with technology-enhanced approaches, such assessment processes can allow separation of ‘teacher’ from ‘evaluator/marker’ roles, enabling the former to help students focus on improving their writing, rather than attempting to serve as both writing guide and grading judge.

References

- ¹ In *Assessing for Learning: Building a Sustainable Commitment Across the Institution* (Sterling, VA, USA: AAHE, 2004), Peggy L. Maki offers a fine introduction to assessment and general overview to validity and reliability (see pp. 92-4). For a helpful discussion of validity in terms of writing assessment, see Roberta Camp’s ‘New Views of Measurement and New Models for Writing Assessment’ (in E. White, W. D. Lutz & S. Kamusikiri’s *Assessment of Writing: Politics, Policies, Practices*, New York: MLA, 1996, pp. 135-147. Edward M. White’s work in the area of reliability, particularly *Teaching and Assessing Writing* (San Francisco: Jossey-Bass, 1994) and *Assigning, Responding, Evaluating: A Writing Teacher’s Guide* (New York: Bedford/St. Martin’s, 1999) continues to provide both a foundation for practice and for subsequent critical debate.
- ² A useful and more general guide to rubrics in multiple academic disciplines is D. D. Stevens and A. J. Levi’s 2005 *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning* (Sterling, VA, USA: Stylus, 2005 [2nd ed. to be published in 2011]).

- ³ For a fuller historical discussion of the initial development of rubrics in large-scale writing assessment, see P. Diederich, J. W. French and S. Carlton's 1961 "Factors in Judgments of Writing Ability" (Princeton, NJ, USA: Educational Testing Service [ETS], RB-61-15), W. McColly's 1970 "What Does Educational Research Say about the Judging of Writing Ability?" (The Journal of Educational Research 64[4], pp. 147-56), and B. Huot, P. O'Neill and C. Moore's 2010 "A Usable Past for Writing Assessment" (College English 72[5], pp. 495-517). The expanding use of rubrics in a range of disciplines and at all educational levels is discussed at length in B. Huot's 2002 (Re)Articulating Writing Assessment (Logan, UT, USA: Utah State UP), B. Broad's 2003 What We Really Value: Beyond Rubrics in Teaching and Assessing Writing (Logan, UT, USA: Utah State UP), and E. White's 2006 Assigning, Responding, Evaluating: A Writing Teacher's Guide (Boston: Bedford/St Martin's, 4th ed.). A counter-argument regarding the development of rubrics in these contexts and their general usefulness in writing assessment is provided by Maja Wilson's 2006 Rethinking Rubrics in Writing Assessment (Portsmouth, NH, USA: Heinemann, cf. pp. 10-26).
- ⁴ Reprinted with permission from Assessing Outcomes and Improving Achievement: Tips and tools for Using Rubrics, edited by Terrel L. Rhodes. Copyright 2010 by the Association of American Colleges and Universities. This and additional rubrics from the AAC&U are available via <http://www.aacu.org/value/rubrics>.
- ⁵ Two valuable discussions of this issue are F. Kemp's "Computers, Innovation, and Resistance in First-Year Composition Programs" and R. E. Miller and M. J. Cripps "Minimum Qualifications: Who Should Teach First-Year Writing?" (both in S. J. McGee and C. Handa's *Discord & Direction: The Postmodern Writing Program Administrator*, Logan, UT, USA: Utah State UP, pp. 105-139).
- ⁶ See S. H. McLeod and M. Soven's *Writing Across the Curriculum: A Guide to Developing Programs* (Newberry Park, CA, USA: Sage Publications, 2000, pp. 94-104; available free online at <<http://wac.colostate.edu/books/>>) and S. Ramey, L. VandeVusse and MaryBeth Gosline's "Using a Written Communication Rubric to Improve Student Writing" (International Journal of Learning 13(10), pp. 67-74) for additional information on creating and sustaining these dialogues between teachers.
- ⁷ R. S. Anderson and J. B. Puckett's 2003 "Assessing Students' Problem-Solving Assignments" (New Directions for Teaching and Learning 95, pp. 81-87) provides a number of examples of rubrics with specific learning outcome foci.



This work is published under the [Creative Commons 3.0 New Zealand Attribution Non-commercial Share Alike Licence \(BY-NC-SA\)](https://creativecommons.org/licenses/by-nc-sa/3.0/). Under this licence you are free to copy, distribute, display and perform the work as well as to remix, tweak, and build upon this work noncommercially, as long as you credit the author/s and license your new creations under the identical terms.