Assessment rubrics: towards clearer and more replicable design, research and practice

Phillip Dawson

Centre for Research in Assessment and Digital Learning, Deakin University, Melbourne, Australia

'Rubric' is a term with a variety of meanings. As the use of rubrics has increased both in research and practice, the term has come to represent divergent practices. These range from secret scoring sheets held by teachers to holistic student-developed articulations of quality. Rubrics are evaluated, mandated, embraced and resisted based on often imprecise and inconsistent understandings of the term. This paper provides a synthesis of the diversity of rubrics, and a framework for researchers and practitioners to be clearer about what they mean when they say 'rubric'. Fourteen design elements or decision points are identified that make one rubric different from another. This framework subsumes previous attempts to categorise rubrics, and should provide more precision to rubric discussions and debate, as well as supporting more replicable research and practice.

Keywords: rubrics; rubric design; assessment design; research synthesis; replicable research

Introduction

The term 'rubric' has varying meanings, and evokes a range of responses from educators (Popham, 1997). A secret scoring sheet, held by the teacher, and only revealed after student work has been marked, is a 'rubric'. An articulation of the standards expected of essays in a particular faculty or department is a 'rubric'. Rubrics can contain detailed grading logic, with numbers and even formulae; alternatively they can have no numbers, and be suggestive of broad quality levels (Sadler, 2009a). One rubric may use generic quality words (e.g. 'good' or 'below standard') whereas another rubric may explain in detail what quality looks like. Some rubrics eschew words in favor of graphics, ranging from emoji to samples of what work should look like for a particular criterion at a particular standard. Since the beginning of its use in education, 'rubric' has not been a particularly clear term:

> 'A couple of decades ago, rubric began to take on a new meaning among educators. Measurement specialists who scored students' written compositions began to use the term to describe the rules that guided their scoring. They could have easily employed a more readily comprehensible descriptor, such as scoring guide, but scoring guide lacked adequate opacity. Rubric was a decisively more opaque, hence technically attractive, descriptor.'
>
> (Popham, 1997, p. 72)

Opacity has not hindered the proliferation of research and writing around rubrics, particularly in the last two decades. Prior to 1997 the term 'assessment rubric' had been used in 83 scholarly papers and in 106 books (Google Books Ngram Viewer, 2015; Google Scholar, 2015). Since then, the term has steadily grown in use, and is

now mentioned in hundreds of papers each year; Figure 1 shows the accumulation of research about rubrics since 1997:

<Figure 1 about here>

In 1997 the 100th paper mentioned 'assessment rubrics'; in 2005 the 1000th paper mentioned the term, and in some time in 2013, the 5000th paper mentioning rubrics was published. A similar trend has existed in books, although data are only available up to 2008.

To provide clarity throught the opacity of the term, Popham's (1997) seminal paper proposed that a rubric must have: evaluative criteria, quality definitions for those criteria at particular levels, and a scoring strategy. Empirical work later in the 1990s found that few teachers shared Popham's understanding, and in practice the term had a constellation of meanings (Wenzlaff, Fager, & Coleman, 1999).

The problem with the combination of conflated meanings and rapid proliferation of 'rubric' is that it has been mandated, evaluated and built into technology, often as if there was some sort of shared understanding. Some departments or institutions have adopted policies that mandate the use of a 'rubric' without providing a working definition of the term, leaving it open to a very diverse array of interpretations that may not be in the spirit the policy makers intended. Researchers have conducted a range of evaluation studies about rubrics, but do not always give operational definitions of what the rubric looked like or how it was used. Learning management systems provide 'rubric' tools that each represent particular interpretations. Against this proliferation, everyday educators have developed enthusiasm or resistance towards rubrics based on what the term has come to mean in their context. This sort of 'label naivety' (Pawson, 2006) is not particular to education, and is common to other social science phenomena as well.

This article aims to provide a language to discuss rubrics. Rather than seek a homogenous definition for the term 'rubric', it provides a framework to map out the heterogeneity of potential rubric interventions. Some work has already been undertaken to provide a language to discuss rubrics. Jonsson and Svingby's (2007) review of 75 studies identifies two classification dimensions: whether a rubric is analytic or holistic, and if it is generic or task-specific. Further distinctions can be drawn between teacher-created rubrics and those that are co-created with students (Reddy & Andrade, 2010). Combining these terms provides greater specificity. An institutional policy might mandate holistic, task-specific, co-created rubrics. A technology tool could be developed to support analytic, task-specific, teacher-created rubrics. A research study might compare holistic, generic, teacher-created rubrics with holistic, task-specific, teacher-created rubrics. As an educator, I feel comfortable with holistic, task-specific, teacher-created rubrics, but less comfortable with other types. This article proposes that rubrics vary by at least 14 such dimensions, and provides a common language to discuss different types of rubrics.

As this article's focus is on making the opacity of rubrics transparent, some preliminary scoping and definition issues must be attended to. For the purposes of this article:

- A *rubric* is a tool used in the process of assessing student work that usually includes Popham's (1997) three essential features: evaluative criteria, quality definitions for those criteria at particular levels, and a scoring strategy.

- A *design element* is a particular variable, choice or dimension that makes one sort of rubric different to another, for example, the *specificity* element is concerned with the differences between task-specific and generic rubrics
- The *framework* proposed in this article is the combination of all of the design elements.

The entire rubric intervention is inside the scope of this framework, not just the physical artifact; for example, the decision to share a rubric with students is of interest and is included in the framework, even though this may not be written on the rubric. This article is limited in scope to the use of rubrics in the assessment process, and includes self, peer, teacher, and exemplar assessment for formative and summative purposes. It does not consider rubric use for other purposes, such as the evaluation of teaching or diagnosis of disease.

Methodology

In this article I take the same approach as Dawson's (2014) synthesis of the design space of mentoring interventions. Just like 'rubrics', 'mentoring' has a diverse range of definitions and understandings, and Dawson used the existing literature to build a framework to express that diversity. This requires a sort of synthesis by configuration approach (Sandelowski, Voils, Leeman, & Crandell, 2012), in that it involves reading a broad set of literature and arranging that literature to create a coherent whole. Importantly while this approach can capture diversity it does not allow this paper to make claims of prevalence or efficacy with respect to particular approaches to using rubrics. In line with that limitation this paper expresses no support or preference for any particular decisions against any design element. The methodology used here does not also make any claims of completeness; the design elements are presented as a starting point rather than as a complete set.

In brief, the strategy used to identify relevant literature involved:

- Reading relevant review studies (e.g. Jonsson & Svingby, 2007; Reddy & Andrade, 2010) and consulting the references of those studies
- Keyword searches in relevant journals, most notably *Assessment & Evaluation in Higher Education; Assessment in Education: Principles, Policy & Practice; Practical Assessment, Research & Evaluation*
- Consultation with rubric researchers face-to-face and via social media

In addition to the design space synthesis approach taken by Dawson (2014) this paper also iteratively tested the framework against the first 100 rubrics captured by the Google Images search 'assessment rubric' (Google Images, 2015). This process provided a valuable reality check for completeness and utility of the framework.

Sample rubric

A sample rubric is provided to offer a reference point for the discussion of the framework. This rubric is taken from a postgraduate education subject and was used to assess a task that required the students to get feedback about their teaching. This rubric is not meant to represent 'good practice', or even usual practice; I have selected it from my own teaching as it is a useful rubric to illustrate particular design elements. I have changed it somewhat to better illustrate the framework.

An excerpt from the task description summary is provided below, and the rubric is provided as Figure 2.

Assignment 2: What you need to do

- *Describe your course context to us*
- *Make a plan to get feedback about your unit/teaching*
- *Gather feedback and analyse it*
- *Make a plan for action, identifying at least two changes to your unit/teaching*
- *Connect with relevant research literature on learning and teaching, ethics, feedback, assessment, and other topics as appropriate*
- *Self-assess your work using the attached rubric and include it in your submission*

<Figure 2 about here>

Fourteen Rubric design elements in brief

An overview of the 14 rubric design elements is provided in Table 1 below. This table includes a concise definition of each design element, references to articles that identify, focus on or mention that element, and how that element is present or absent in the sample rubric.

<Table 1 about here>

Fourteen Rubric design elements in detail

In this section the 14 design elements in Table 1 are defined and discussed with reference to the literature and the example rubric. Some elements are the specific topic of study in published literature, such as Sadler's (2009a) discussion of analytic vs holistic rubrics. Other elements are discussed with respect to literature that mentions but does not focus on that specific element.

Specificity: the particular object of assessment

Rubrics are sometimes discussed as being 'generic' or 'specific' (e.g. Tierney & Simon, 2004). However the specificity of rubrics is more than an either-or binary. The example provided is a *task-specific* rubric, as it applies to a specific instance of assessment in a particular course unit. At a more generic level, *task-type rubrics* have been developed for reuse across a range of individual tasks, such as Timmerman et al's (2010) rubric for assessing 'scientific writing'. At an even broader level *department-wide* or *institution-wide rubrics* exist to assess all work across entire departments.

Secrecy: who the rubric is shared with, and when it is shared

Some rubrics are never shared with students, and exist as a secret scoring sheet to assist markers. Others are shared as a part of marking and feedback provision. There has been substantial debate around the sharing of criteria and rubrics with students (e.g. Torrance, 2007). The example rubric was shared with students at the start of semester as part of the package provided to specify the assessments.

Exemplars: work samples provided to illustrate quality

While not a mandatory element, many uses of rubrics incorporate exemplars of work that demonstrate particular criteria or quality descriptions. Examples of student work for a particular task can be useful if a less specific rubric is used (Tierney & Simon, 2004). The example rubric was provided to students with an exemplar that had been through the same marking and feedback process the students' work was to undergo.

Scoring strategy: procedures used to arrive at marks and grades

The element *scoring strategy* has been taken from Popham's (1997) three essential features. In line with earlier work on assessment judgements (e.g. Sadler, 1989), Popham discusses analytic and holistic scoring strategies. Holistic scoring strategies require the user to take all of the criteria into consideration and aggregate them to a single overall quality judgement. Analytic scoring strategies require the user to make a series of judgements about individual criteria, which are often combined formulaically to produce an overall mark. For a discussion of the pros and cons of analytic and holistic scoring, see Sadler (2009a).

Scoring strategies have advanced somewhat since Popham's definition. Rather than being a simple addition of scores, analytic approaches can incorporate complex logic (Sadler, 2009a). Modern scoring strategies can also involve multiple scorers, and in these cases it can be necessary to include procedures to resolve disputes (Johnson, Penny, & Gordon, 2000). Software can be used to make some or all of the judgements on a rubric (Dimopoulos, Petropoulou, & Retalis, 2013). The scoring strategy on the sample rubric is analytic and requires the user to consider some minimum requirements for particular levels. It is also influenced by faculty policy which required a second independent assessment of all 'Fail' grades.

Evaluative criteria: overall attributes required of the student

Evaluative criteria are 'used to distinguish acceptable responses from unacceptable responses' (Popham, 1997, p. 72). These typically occupy the leftmost column on a rubric. Examples of evaluative criteria include 'organization, mechanics, word choice and supporting details' (Popham, 1997, p. 72), and 'introduction, argument, summary and conclusions, and references' (Biggs & Tang, 2007, p. 210). Although evaluative criteria may be considered an essential element (Popham, 1997), in many rubrics available online, and in the sample rubric, they are missing.

Quality levels: the number and type of levels of quality

When a rubric is represented in a table, quality levels usually occupy the top row. The words used to identify quality levels include learning outcomes taxonomies like Bloom's or SOLO, grade levels, or statements of student proficiency (e.g. Timmerman et al., 2010). It is possible to combine these, for example Biggs and Tang (2007, p. 210) provide a rubric that uses A-D grade descriptors as its headings, which map to the four levels of the SOLO taxonomy. In contrast, Fluckiger (2010) argues for 'single-point' rubrics, which only illustrate the acceptable level of performance. There is some debate as to whether all criteria need representation at all quality levels (Sadler, 2009b; Tierney & Simon, 2004). In the sample rubric the quality levels correspond to the grade scheme of the faculty.

Quality definitions: explanations of attributes of different levels of quality

Another of Popham's (1997) essential features of a rubric, quality definitions are also called *quality descriptors* or (somewhat confusingly) *criteria* (Sadler, 2009b). When a rubric is shown as a table, each quality definition typically occupies one cell and represents a particular *evaluative criterion* at a particular *quality level*. Rubric users rely on these to inform judgements about quality, and they are often used as a way to explain what a particular evaluative criterion looks like at a particular level. There is substantial variation between rubrics in the level of detail provided in quality definitions. In the sample rubric, quality definitions contain moderate detail and address different attributes across quality levels.

Judgement complexity: the evaluative expertise required of users of the rubric

Identifying the quality definitions that are represented in a piece of student work requires the rubric user to exercise judgement. This is related to, but not the same as the rubric's scoring strategy. Consider a rubric for an essay task with an analytic scoring strategy, and a range of evaluative criteria. One of these criteria, 'clarity of expression' has the following quality definitions: 'bad'; 'acceptable'; and 'good'. Distinguishing between these three levels requires making a 'qualitative judgement': 'one made directly by a person, the person's brain being both the source and the instrument for appraisal' (Sadler, 2009b, p. 47). These judgements are necessarily complex and expert.

The sample rubric requires some qualitative judgements, but also some judgements that are more 'analytic' (Sadler, 2009b), with a focus on structure or the presence of particular information. These judgements are less complex and require less expertise; some analytic judgements on rubrics can even be made automatically by computers (Dimopoulos et al., 2013).

Users and uses: who makes use of the rubric, and to what end

Rubrics are more than just a tool used to support assessors in making summative judgements. Teachers also use rubrics as a way to provide feedback information (Nordrum, Evans, & Gustafsson, 2013). It is important to note that teachers are not a homogeneous group, and in the modern university marking is often devolved to a group of 'para academic specialists' rather than traditional academics (Macfarlane, 2011). Students can use rubrics in a range of ways, including self- and peer-assessment, and in interrogating the requirements of a task (Andrade & Du, 2005; Panadero & Romero, 2014). Students use the sample rubric for: planning their response to the task; in-class formative peer assessment; and self-assessment. Teachers use the sample rubric to provide summative grading and feedback information.

Creators: the designers of the rubric

The creation of rubrics is not exclusively the domain of the teacher setting the assessment. Textbook publishers have, for decades, been providing assessment rubrics (Popham, 1997); institutions and professional bodies mandate some rubrics; researchers validate and publish rubrics (Timmerman et al., 2010); and online sources provide extensive rubric banks from a variety of creators (Dornisch & McLoughlin, 2006). With the rise of para-academic specialists (Macfarlane, 2011), rubric experts may be called upon in the same way researchers call on statistical consultants.

Students can also be involved in the creation of rubrics and identification of criteria (Andrade & Du, 2005; Boud & Soler, 2015). As the teacher in charge of the course unit, I created the sample rubric.

Quality processes: approaches to ensure the reliability and validity of the rubric

Various processes can be used to validate a rubric or test its reliability. Timmerman et al (2010) provide an example of the use of quality processes throughout the design, refinement and implementation of a rubric. They conducted statistical tests to ensure the rubric was reliable, in as far as multiple markers would arrive at similar marks for the same piece of work. For a review of possible approaches to inter-marker reliability, see Johnson et al (2000). Timmerman et al (2010) took multiple approaches to ensuring the validity of their rubric: consulting existing rubrics; comparing to authentic criteria used to judge similar tasks; consulting pedagogical experts; and iterative development and feedback from stakeholders who were also content experts. In contrast, no formal quality processes were applied to the sample rubric, although it was iteratively refined based on student feedback and performance each semester.

Accompanying feedback information: comments, annotation, or other notes on student performance

Rubrics can be a tool to articulate feedback information, and are sometimes used instead of in-text commentary or as a supplement (Nordrum et al., 2013). The sample rubric is accompanied by several pieces of feedback information. When used as a formative in-class tool, students use the rubric as stimulus for a feedback discussion. When used summatively, a narrative is provided by the marker about the differences in judgement between their judgement and the student's self-assessment, along with in-text comments.

Presentation: how the information in the rubric is displayed

Definitions of rubrics tend to mention presentation elements, and usually describe a grid, table or matrix populated by text (e.g. Sadler, 2009a). Examples of rubrics in the literature tend to conform to this norm (e.g. Timmerman et al., 2010). However in practice the information on rubrics can be presented in other forms. When reviewing samples of rubrics for this paper (Google Images, 2015) the most common non-text representation was the use of images in quality definitions, such as samples of what work might visually look like for a given criterion, or various smiley faces. Outside of the domain of education, the Bristol Stool Chart (Lewis & Heaton, 1997) provides an example of an instrument that meets the definition of a rubric but relies heavily on drawings to show the qualities of different feces. It is possible to conceive of other representations, such as Venn diagrams or sophisticated technology-mediated rubrics, which may communicate the same information but provide additional affordances. The sample rubric is a paper-based table of text.

Explanation: instructions or other additional information provided to users

Rubrics are accompanied by varying explanations, however studies about the use or efficacy of rubrics often omit these explanations from publication. Sometimes these instructions are as simple as being asked to complete the rubric (Hafner & Hafner, 2003); in some other cases more detailed instructions are provided (Panadero &

Romero, 2014). The sample rubric was accompanied by the statement 'Use this to self- and peer-assess. Submit a highlighted self-assessed copy.' It was also supported with a peer-facilitated discussion in class where students interrogated the rubric themselves.

Discussion and conclusions

The term 'rubric' has been the site of conflation and confusion since at least the 1990s (Wenzlaff et al., 1999). This article's framework of 14 design elements demonstrates the breadth of assessment practice undertaken under the banner of that one word. In doing so, it subsumes previous attempts at taxonomies, and identifies new variations. In this section, potential benefits and limitations of the framework are discussed for researchers, practitioners, software manufacturers and policy makers.

Researchers may benefit from the framework as a way to clearly and concisely communicate what a particular rubric intervention entailed. In the absence of this structure, rubric research has proliferated that does not provide essential information. Research that reports rigorous evaluations of rubrics might not state if the rubric is secret or shared; replacing feedback or complementing it; or how the rubric was explained to its users. Although many published studies include the rubric in their body or appendix (e.g. Tierney & Simon, 2004; Timmerman et al., 2010), several of the design elements presented in this paper are not apparent from those rubric documents in isolation. These papers confirm something called a 'rubric' is a good idea, but leave some gaps about what exactly that something is.

Compared to other disciplines, educational researchers publish few replication studies (Makel & Plucker, 2014). Replicable research requires sufficient detail on the research method, which in educational research includes the particular intervention. None of the rubric studies referenced throughout this article would be candidates for a replication study, in large part due to insufficient information about the rubric intervention employed. However if combined with sufficiently detailed and robust research design, the rubric framework in this article could provide sufficient information about the intervention design for a literal or conceptual replication study. This would also provide practitioners with enough information to implement a particular design in their own teaching.

Practitioners designing a rubric could also fruitfully reflect on the design elements as a way of revealing any 'unknown unknowns' or assumptions about rubrics. This could be particularly useful in bringing course teams towards a shared understanding, and helping them to avoid the confusion discussed by Wenzlaff et al (1999).

Since rubrics have become required by policy in various settings, this framework may provide policy makers with an opportunity to identify the particular practices they are mandating. Secret teacher-created generic analytic rubrics and shared co-created task-specific holistic rubrics serve very different purposes, and clarification in policy about what sort of 'rubric' is required in a particular setting would help achieve the desired result.

Computer software manufacturers could benefit from reflecting on the design elements in this framework, both at design time and when marketing the capabilities of their tools. This framework subsumes the elements present in software-centric rubric taxonomies like the IMS Global Learning Consortium (2005) rubric specification, and could be used to specify the pedagogical use cases of a rubric tool.

This framework provides no guidance on the pros and cons of particular decisions made about individual design elements. But work has already been undertaken about some elements, for example the design of a scoring strategy could be informed by Johnson et al's (2000) work on resolving disputes between assessors. The framework could provide a useful structure for a synthesis of existing literature on the effect of rubrics on learning. It could also help identify worthwhile opportunities for new empirical work.

This paper has been limited in scope to the use of rubrics to assess students' work, but rubrics are also used for other purposes. Rubrics exist across a range of diverse contexts, for example, peer review of teaching (Magno, 2012); fecal classification (Lewis & Heaton, 1997); and evaluating research (Wong, Westhorp, Pawson, & Greenhalgh, 2013). The framework in this article may provide transparency for the opaque term 'rubric' in other domains as well.

Notes on contributors

Phillip Dawson is Associate Professor and Associate Director of the Centre for Research in Assessment and Digital Learning at Deakin University. Email: p.dawson@deakin.edu.au

References

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation, 10*(3), 1-11.

Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university*. Maidenhead, Berkshire, England: Open University Press.

Boud, D., & Soler, R. (2015). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*. doi: 10.1080/02602938.2015.1018133

Dawson, P. (2014). Beyond a Definition: Toward a Framework for Designing and Specifying Mentoring Models. *Educational Researcher, 43*(3), 137-145. doi: 10.3102/0013189x14528751

Dimopoulos, I., Petropoulou, O., & Retalis, S. (2013). *Assessing students' performance using the learning analytics enriched rubrics*. Paper presented at the Proceedings of the Third International Conference on Learning Analytics and Knowledge, Leuven, Belgium.

Dornisch, M. M., & McLoughlin, A. S. (2006). Limitations of web-based rubric resources: Addressing the challenges. *Practical Assessment, Research & Evaluation, 11*(3), 1-8.

Fluckiger, J. (2010). Single point rubric: a tool for responsible student self-assessment. *Delta Kappa Gamma Bulletin, 76*(4), 18-25.

Google Books Ngram Viewer. (2015). Google Ngram Viewer. Retrieved 28 April, 2015, from https://books.google.com/ngrams/graph?content=assessment+rubric&year_start=1800&year_end=2000&corpus=15&smoothing=3&share=&direct_url=t1%3B%2Cassessment rubric%3B%2Cc0

Google Images. (2015). assessment rubric - Google Search. Retrieved 28 April, 2015, from https://www.google.com.au/search?q=assessment+rubric&espv=2&biw=1280

&bih=1255&source=lnms&tbm=isch&sa=X&ved=0CAYQ_AUoAWoVChM
I__mRkJKMxwIVxWGmCh010ARN

Google Scholar. (2015). "assessment rubric" - Google Scholar.   Retrieved 28 April, 2015, from https://scholar.google.com.au/scholar?q="

Hafner, J., & Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer‑group rating. *International Journal of Science Education, 25*(12), 1509-1528. doi: 10.1080/0950069022000038268

IMS Global Learning Consortium. (2005, 2 June 2005). IMS Rubric Specification. Version 1.0 Final Specification. from http://www.imsglobal.org/ep/epv1p0/imsrubric_specv1p0.html

Johnson, R. L., Penny, J., & Gordon, B. (2000). The Relation Between Score Resolution Methods and Interrater Reliability: An Empirical Study of an Analytic Scoring Rubric. *Applied Measurement in Education, 13*(2), 121-138. doi: 10.1207/S15324818AME1302_1

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130-144. doi: http://dx.doi.org/10.1016/j.edurev.2007.05.002

Lewis, S. J., & Heaton, K. W. (1997). Stool Form Scale as a Useful Guide to Intestinal Transit Time. *Scandinavian Journal of Gastroenterology, 32*(9), 920-924. doi: 10.3109/00365529709011203

Macfarlane, B. (2011). The Morphing of Academic Practice: Unbundling and the Rise of the Para-academic. *Higher Education Quarterly, 65*(1), 59-73. doi: 10.1111/j.1468-2273.2010.00467.x

Magno, C. (2012). Assessing higher education teachers through peer assistance and review. *The International Journal of Educational and Psychological Assessment, 9*(2), 104-120.

Makel, M. C., & Plucker, J. A. (2014). Facts Are More Important Than Novelty: Replication in the Education Sciences. *Educational Researcher, 43*(6), 304-316. doi: 10.3102/0013189x14545513

Nordrum, L., Evans, K., & Gustafsson, M. (2013). Comparing student learning experiences of in-text commentary and rubric-articulated feedback: strategies for formative assessment. *Assessment & Evaluation in Higher Education, 38*(8), 919-940. doi: 10.1080/02602938.2012.758229

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice, 21*(2), 133-148. doi: 10.1080/0969594X.2013.877872

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. London, England: Sage.

Popham, W. J. (1997). What's wrong - and what's right - with rubrics. *Educational Leadership*(2), 72.

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education, 35*(4), 435-448. doi: 10.1080/02602930902862859

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144. doi: 10.1007/BF00117714

Sadler, D. R. (2009a). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education, 34*(2), 159-179. doi: 10.1080/02602930801956059

Sadler, D. R. (2009b). Transforming Holistic Assessment and Grading into a Vehicle for Complex Learning. In G. Joughin (Ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 1-19): Springer Netherlands.

Sandelowski, M., Voils, C. I., Leeman, J., & Crandell, J. L. (2012). Mapping the Mixed Methods–Mixed Research Synthesis Terrain. *Journal of Mixed Methods Research, 6*(4), 317-331. doi: 10.1177/1558689811427913

Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation, 9*(2), 1-10.

Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2010). Development of a 'universal' rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education, 36*(5), 509-547. doi: 10.1080/02602930903540991

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post‐secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice, 14*(3), 281-294. doi: 10.1080/09695940701591867

Wenzlaff, T. L., Fager, J. J., & Coleman, M. J. (1999). What is a Rubric? Do Practitioners and the Literature Agree? *Contemporary Education, 70*(4), 41.

Wong, G., Westhorp, G., Pawson, R., & Greenhalgh, T. (2013). Realist Synthesis: RAMESES Training Materials. Retrieved 26 January, 2015, from http://www.ramesesproject.org/index.php?pr=Project_outputs
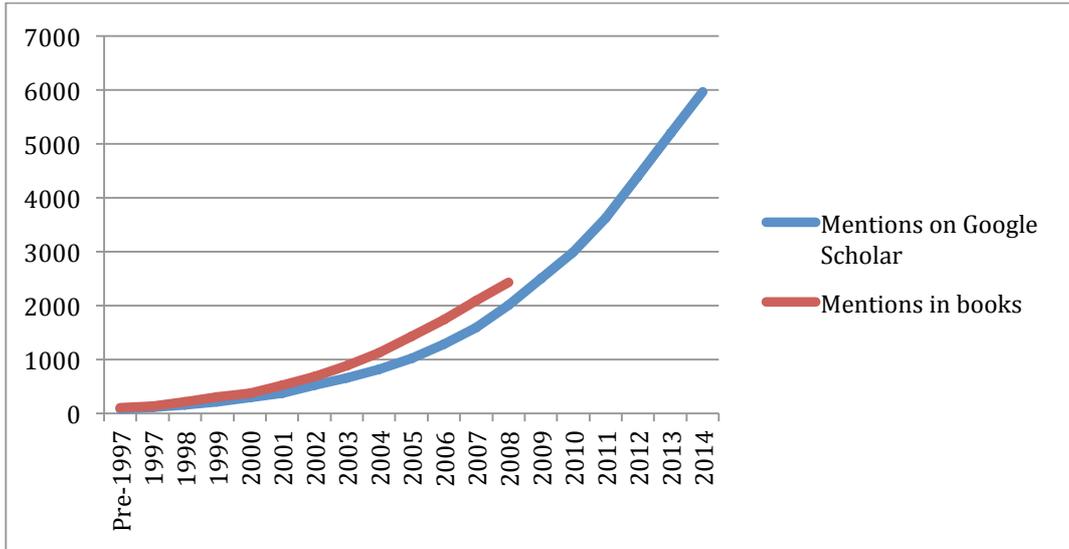
Figure 1: Cumulative frequency of Google Scholar and Google Books items that used the term 'assessment rubric' in their title, abstract or body. Data sourced April 28, 2015.

Fail	Pass	Credit	Distinction	High Distinction
	None of the Fail items, and all of the following	None of the Fail items, all of the Pass items and most of the following	None of the Fail items, all of the Pass items, most of the Credit items, evidence of peer review of teaching, and most of the following	None of the Fail items, all of the Pass items, all of the Credit items, most of the distinction items, and most of the following
Insufficient data about teaching context to understand the rest of the assignment	Collected feedback from students	Evidence of peer review by a colleague	Three or more sources of feedback	Four or more sources of feedback
Parts of the assignment are missing	Clear documentation of feedback collection process	Plan for action is well supported by scholarly literature	Mixture of qualitative and quantitative feedback	Feedback provides input on student learning, teaching, and the student experience
Did not collect or source any feedback from students	Unambiguous connection between feedback and plan	Substantial use of scholarly literature, including >=6 sources	Consideration of the ethics of feedback about teaching	Particularly rigorous analysis methods
Major ethical problems with data collection	Targeted use of >=4 sources	No ethics problems	Connection to research literature about feedback	Connection to the research literature about ethics in feedback about teaching
Any evidence of academic integrity problems		Based on the weight of evidence (feedback, argument), self/peer marker expects action plan will work	Synthesis of literature and feedback into connected themes	Writing, methods and scholarship of similar quality to a higher refereed education conference paper
Very little use of research literature		If research from other education sectors (school/VET) is used, its applicability to higher education is considered carefully	Some critique of feedback instruments or the role of education	
Substantial problems with clarity or referencing			If research from other sectors teaching and learning	
			Use of existing feedback instruments, or clear justification for the development of new instruments	

Figure 2. Sample rubric

Design element	References	Sample rubric
Specificity: the particular object of assessment	Tierney and Simon (2004): generic rubrics vs task-specific Dornisch and McLoughlin (2006): challenges of using non-task-specific rubrics from the web Timmerman et al. (2010): example of a rubric to assess 'scientific writing' in general	Task-specific
Secrecy: who the rubric is shared with, and when it is shared	Torrance (2007): challenges of sharing criteria and different interpretations (not rubric-specific)	Shared with task description
Exemplars: work samples provided to illustrate quality	Tierney and Simon (2004): argues for providing exemplars with rubrics	One example of high-quality work was provided with a completed rubric
Scoring strategy: procedures used to arrive at marks and grades	Sadler (2009a): different types of scoring logic Johnson et al. (2000): score resolution when assessors disagree Popham (1997): rubric definition mentions scoring strategies Dimopoulos et al. (2013): use of computers in a scoring strategy	Analytic. Cumulative scoring logic to arrive at broad grade. Faculty policy required double-marking of fails.
Evaluative criteria: overall attributes required of the student	Popham (1997): rubric definition mentions evaluative criteria	Absent
Quality levels: the number and type of levels of quality	Sadler (2009b): mentions quality levels, noting that they need not be uniform across criteria Fluckiger (2010): provides rationale for using just one quality level Biggs and Tang (2007, p. 210): levels aligned with SOLO	Five levels corresponding to grade descriptors
Quality definitions: explanations of attributes of different levels of quality	Popham (1997): rubric definition mentions quality definitions Sadler (2009b): notes terminology is not uniform around quality descriptors and criteria. Tierney and Simon (2004): encourages consistency across levels	Present but inconsistent attributes across performance levels.
Judgement complexity: the evaluative expertise required of users of the rubric	Sadler (2009b): 'qualitative judgements' vs 'analytic judgements' Dimopoulos et al. (2013): computers making judgements in 'learning analytics enriched rubrics'	Moderate: mixture of analytic and qualitative judgements

Design element	References	Sample rubric
Users and uses: who makes use of the rubric, and to what end	Nordrum et al. (2013): teachers using rubrics to communicate feedback information Panadero and Romero (2014); Andrade and Du (2005): particular student uses of rubrics Dimopoulos et al. (2013): computers as users of rubrics	Teachers use for summative assessment; students use for planning and self-assessment; students use for formative peer assessment
Creators: the designers of the rubric	Andrade and Du (2005); Boud and Soler (2015): rubrics co-created by students and teachers Timmerman et al. (2010): researchers creating a rubric	Teacher
Quality processes: approaches to ensure the reliability and validity of the rubric	Johnson et al. (2000): inter-rater reliability Timmerman et al. (2010): example of rubric that has undergone reliability and validity testing	No formal quality processes. Informal refinement based on student feedback and performance.
Accompanying feedback information: comments, annotation, or other notes on student performance	Nordrum et al. (2013): compared rubric-articulated feedback with in-text commentary	In-class: rubric acts as a stimulus for peer feedback discussion. Summative marking: rubric accompanied by narrative from marker, and in-text comments.
Presentation: how the information in the rubric is displayed	(e.g. Sadler, 2009a): usual presentation is a grid, table or matrix of text Google Images (2015): a range of examples of how rubrics are presented	Paper-based table of text
Explanation: instructions or other additional information provided to users	Hafner and Hafner (2003): provided minimal instruction Panadero and Romero (2014): more detailed instructions	Minimal: 'Use this to self- and peer-assess. Submit a highlighted self-assessed copy.'

Table 1. Summary of the rubric design elements.