

# **Using rubrics in higher education: Some suggestions for heightening validity and effectiveness**

Peter McDowell

*Charles Darwin University*

## **Abstract**

After isolating several intrinsic problems with the generic structure of assessment rubrics, especially in relation to their validity and effectiveness, the paper canvasses an alternative approach that has been trialled successfully with large, diverse cohorts of primarily online students undertaking graduate level, pre-service teacher education—but with much broader application. Requiring only modest, additional preparation, and with various options available for semi-automation, the alternative approach is able to sustain valid, reliable, and efficient assessment of rich, contextually embedded, problem-based learning by teams of independent assessors, with each assessor generating a combination of bespoke, individualised feedback and structured, proximally relevant, feedforward commentary. The essential difference between the two approaches— instantiating pre-populated analytic rubrics versus interleaving qualitatively bounded tiers of dialectically differentiated commentary—is then accounted for, both mathematically and sociologically. The paper concludes with some pedagogical implications and allied recommendations for assessment design.

**Keywords:** rubrics, higher education, assessment, pedagogy, dialectics

## **Introduction**

The use of rubrics has a lengthy tradition, but one that differs significantly from current practices in the design and application of assessment rubrics. Historically, rubrics have denoted the various means of marking, distinguishing, and amplifying textual content through the addition of headings or marginalia, often within a liturgical setting (Popham, 1997); typically, these inscriptions would be made in red lettering, the term rubric being etymologically related to the use of ochre as a writing material (Stevenson, 2010). Now, within a more recent, educational context, assessment rubrics have emerged to form a distinct, structured, educational genre (Martin & Rose, 2008) conveying statements of hierarchically differentiated performance (Goodrich, 1996), with direct (or sometimes implicit) reference to another educational genre: the assignment specification.

Importantly, since the mid-1990s, there has been continued advocacy for the broader use of assessment rubrics in higher education, with calls coming from multiple disciplines (Connelly & Wolf, 2007): that is, not just from within faculties of education (Allen & Tanner, 2006). Nonetheless, within the educational context, assessment rubrics

are frequently used in schooling, particularly secondary schooling (usually as a prerequisite for cross-sector moderation), with coverage now encroaching on junior primary (elementary) schools. Moreover, this progress has seen growing levels of institutionalisation, with assessment rubrics (as a means of ‘objectifying’ assessors’ judgements) becoming embodied within institutional policy statements—and, more significantly, within student expectations: from personal experience, many newly enrolling students will ask immediately, upon receiving a detailed assignment specification: ‘where is the rubric?’

High levels of operationalisation (i.e., the materialisation of intellectual product through the design of systems) are now encountered in major software products, with assessment rubrics (or frameworks for their ready production) typically provided as optional toolsets within digital learning management systems.

Indeed, as with many successful ‘innovations’, assessment rubrics have reached the point of becoming naturalised, as one of many habitual (and, therefore, often unquestioned) practices within higher education.

How, then, to evaluate this situation?

### **Research focus**

Although the uptake within higher education has been broad and rapid, the conceptual basis of assessment rubrics has been inadequately explored. The research literature tends towards advocacy, with the preconditions for valid and effective application not being subjected to rigorous questioning (Reddy & Andrade, 2010). Indeed, the literature contains very limited critique: where present, critique has tended towards purism (i.e., better design execution) rather than reconceptualisation (Baryla, Shelley, & Trainor, 2012).

As a potential remedy, this paper is an initial contribution towards more a thorough, rigorous questioning of assessment rubrics: indeed, as a more radical, constructive critique, framed in relation to an alternative approach that has arisen in response to the emergence of significant deficiencies in the application of assessment rubrics within large, complex settings—deficiencies in terms of both validity and effectiveness.

Overall, the paper will argue that the alternative approach (sketched below) is superior on several grounds: i.e., mathematically, sociologically, and pedagogically.

The counter claim, that rubrics are invalid and ineffective, will not be maintained: rather, that the preconditions for their valid application demand a much more restricted application (i.e., they need to be kept within their theoretical and practical limits; the limits of their validity and effectiveness).

Or, to say this more pointedly, is the continued use of invalid rubrics worth the risk? Or again, pragmatically speaking, is there a sound alternative, with a comparable preparation workload, that can bring superior results during the assessment process, in terms of both heightened validity *and* greater efficiency?

## **Generic structure**

In general terms, assessment rubrics can be considered as a type of psychometric instrument, reflecting (ideally) the professional judgement of impartial, qualified assessors.

There are several distinguishing features.

It is important to note that an instantiated assessment rubric accompanies (or supplements) an entirely separate specification: namely, the assignment, the set task or project, or a statement of the capabilities under examination. In particular, each individual assessment rubric isolates a small, pre-specified set of independent factors, which represent anticipated characteristics of successful performance (see **Fig. 1**). Each factor is further divided into developmental gradations: i.e., a short, discrete sequence of performance strata. Furthermore, each stratum is assigned a generic, objective statement of typical, commensurate performance.

The performance statements within a sequence usually share nominal and verbal elements (in the case of English), and gradation or stratification is usually achieved through adjectival and adverbial modification (Tierney & Simon, 2004), including quantitative and qualitative differentiation, and sometimes polarity to designate non-attainment. Conceptually, this organisation equates to a per-factor spectrum of performance achieved through intensification and attenuation of the target behaviour.

Assessment rubrics are often arranged in tabular form, with the effect that the strata become co-aligned across factors: i.e., they map by ordinal position. The tabular format (being algorithmically suggestive) encourages the development of scoring rubrics (Moskal, 2000), where factors and tiers (of strata) are weighted and then aggregated (usually as a linear combination, potentially summing to unity).

Assessment rubrics are typically distributed separately from the assessment items (forming, *de facto*, a supplementary specification), and various claims have been made for improving the transparency of assessment, increasing the clarity of expectations (Goodrich Andrade, 2000), and in promoting students' self-assessment (Andrade, 2007).

## **Intrinsic problems**

The preconditions for valid usage are not well understood, and are usually overlooked at the point of application.

What are some of these preconditions?

Perhaps most importantly, scoring rubrics with pre-weighted factors and strata need adequate calibration before use, a non-trivial pre-assessment task, which (in practice) tends to promote the over-solidification of assessment items, thus heightening the risk of students' academic misconduct—the recycling of assignments and their responses.

Such calibration, of course, is standard practice in academic research (i.e, trialling, configuring, and refining psychometric instruments before broad-scale administration),

	High (Weighted)	Medium (Weighted)	Low (Weighted)	None (Weighted)
Factor 1 (Weighted)	Statement of performance	Statement of performance	Statement of performance	Statement of performance
Factor 2 (Weighted)	Statement of performance	Statement of performance	Statement of performance	Statement of performance
Factor 3 (Weighted)	Statement of performance	Statement of performance	Statement of performance	Statement of performance
Overall (Aggregate)	General commentary			

**Fig. 1.** A generic assessment rubric (with optional scoring)

but this knowledge hasn't transferred readily to the teaching domain. Indeed, this scenario supports a broader thesis concerning the general *non-transferability* of endemic, disciplinary-based practices: in this case, the conceptual basis is psychology, with the disciplinary boundaries of cognates (such as education) largely impervious to the successful transmission of scientific praxis. Accordingly, deliberate interdisciplinary intervention is required to re-ground the desired practice with an alternative context and discourse, such as is occurring in this paper. (Several disciplines are active here: education, linguistics, mathematics, sociology, and philosophy.)

An assumption of contextual consistency (or commensurability), which is needed to apply the same rubric to the same cohort, usually won't hold for rich assignments and tasks requiring individuation (whether affected by setting or personal requirements).

Tiered, factor-based pre-specification violates the constructivist principle (Oxford, 1997): an openness to creative, unanticipated performance. Indeed, factors are often not orthogonal (i.e., they are variously interdependent, perhaps to uncertain degrees). Moreover, the co-alignment of ordinal strata across factors is not accounted for in the design, which assumes that the separate factors share the same rates of intensification and attenuation: on what basis is there a single rate? Factors can often be combined in different ratios to achieve similarly valid outcomes, undermining the assumption of contextually pervasive orthogonality.

The next problem is very significant for students. Presenting the full developmental spectrum for each factor violates the principle of proximal development (Poehner, 2012): many students won't make adequate sense of statements concerning distal (higher) performance, which easily leads to confusion, reduction, and the projection of existing attitudes.

In terms of the assessment process, administration of the instrument requires reliable identification (from the submission) of an applicable stratum for each pre-identified factor: this step is not fault-tolerant, and potentially unreliable. Moreover, with scoring rubrics, the factor-based scores aggregate into an overall grade, a bottom-up approach,

which is ineffective for grade-based certification, as it requires additional grade-based calibration.

When interacting with students, instructors' pre-submission and post-assessment distribution of the very same performance statements (embodied in the rubric) conflates the different purposes of assessor feedback and feedforward advice. Indeed, the rubric (as a document), if pre-distributed, acts as a supplementary specification, with high potential to mislead or misdirect students through its compartmentalised, reductive, and linear structures: there is an essential gap between post-performance analysis and the synthesis of capabilities needed for *in vivo* execution.

### **An alternative approach**

Is there an alternative approach that avoids many of these problems (particularly those relating to validity), yet retains most of the benefits of assessment rubrics?

The following approach (**Fig. 2** refers) is offered as a potential candidate, and incorporates these principal design characteristics:

- encourage students' canvassing of possible valid responses (both likely and unlikely), tolerating approximation and partially competent articulation;
- encourage pre-submission discussion of proximal improvement (e.g., through the narration of progressive attainments and strategic sharing);
- as feedback, report on meaningful clusters or 'constellations' of disciplinary practices needed to complete the assessment task;
- as feedforward commentary, advise on feasible, proximal improvements only (cf. Lantolf, & Poehner, 2004);
- differentiate proximal performance nominally and verbally: i.e., neighbouring performance statements should be conceptually discontinuous, ensuring that they are neither quantitatively nor qualitatively commensurable—without this condition there is no disjunct, no spur to reconsider the earlier approach taken;
- certify the totality: i.e., grade at a coarse level only, thus producing a statement about the overall validity of each student's response (within broad grade bands, three levels can usually be managed reliably: e.g., C-, C, and C+);
- partial deviations from the gross attainment can be used to temper the grade allocation, without having to reconsider or abandon any preset weighting scheme (e.g., episodic conceptual weakness might shift a C- to a D+);
- any single, pre-identified constellation of performance characteristics need not be invoked: it is permissible for a response to underplay or even omit a particular sphere of performance (and still remain valid); and
- no pre-written statement of performance need be invoked: there is an option for bypassing or overriding preconceived expectations for any particular constellation.

	Actual	Proximal
Constellation 1 (non-mandatory)	Statement of performance	Advice for improvement
Constellation 2 (non-mandatory)	Statement of performance	Advice for improvement
Constellation 3 (non-mandatory)	Statement of performance	Advice for improvement
Overall (Threshold)	Bespoke commentary	

**Fig. 2.** An alternative approach

### **Additional preparation**

This alternative approach typically requires the following, additional preparatory activities (as compared with those needed to develop a successful rubric):

- locate the criteria necessary for the allocation of coarse grades: usually, grade-level characteristics will be specified institutionally; otherwise, identify the threshold criteria for competent performance;
- decide on the priority areas of focus for assessment feedback or feedforward advice;
- identify the main, non-exhaustive, non-mandatory suites of typical performance characteristics (previously termed ‘constellations’) whose integral (or synthetic) fulfilment provides clear evidence towards the grade assignment and the feedback/feedforward focus;
- anticipate a range of potential responses (varying in disciplinary sophistication) for each element of the assignment task (identifiable from directives and imperatives);
- map the constellations of disciplinary performance (or conceptual sophistication) needed to produce these responses onto meaningful ordinal categories (recommend  $k \leq 4$ ): threshold performance is either at either positions one or two (position zero being reserved for non-performance);
- for each constellation, prepare statements that substantially differentiate the ordinal categories (i.e., distinct tiers of performance), taking care to address actual performance together with practical advice (or, alternatively, interrogative prompting) on attaining the next (proximal) tier of performance;
- the key to this last step is to use ensure that actual *versus* proximal performance levels are distinguished categorically: i.e., in English, nominally and verbally through the use of different noun/verb combinations; indeed, with some care and experience, a progressive chain of suitable noun/verb combinations can often be derived dialectically (n.b., this is a non-trivial task, requiring domain expertise and proficiency in dialectics); and

- group and then partially order the salient constellations according to the anticipated hyperthematic structures (Martin, 1992) of the generic texts (or work products) that you have encouraged students to produce through the design of your assignment and its written specification: in other words, always consider your workflow.

Note that most of these steps are amenable to operationalisation, allowing for semi-automation of the assessment process: e.g., the performance statements (coupling actual and proximal performance) can be added to a comment bank prior to receiving students' submissions.

### **Trialling**

This alternative approach has been developed, trialled, and fully implemented in complex, semestral higher education modules delivered biannually since 2012.

The teaching context comprises graduate entrants undertaking postgraduate teacher education in cross-curricular capabilities: literacy and numeracy, with an emphasis on pedagogy and multi-disciplinary coherence. Usually, individual students find at least one of these areas to be very challenging, reflecting disciplinary biases and linguistic backgrounds (which also include substantial numbers of ESL and EFL learners). The student cohorts are large, ranging between 200 and 800 enrolments per module (on a steady, upward trajectory). The delivery modes are predominantly external or blended.

Students typically undertake complex, problem-based assignments with substantial, multiple design elements and expectations for personal contextualisation, two aspects which students often find difficult to combine. (The contexts are multi-jurisdictional.)

Assessment is conducted by multiple markers (small teams, up to four) with mixed levels of experience in the teaching domain and with the marking approach. Various institutional protocols determine a need for ongoing cross-marker moderation, as well as expectations to explain patterns of performance (criterion rather than norm-referenced), particularly in relation to deviations from historical grade distributions. Non-completions are monitored carefully.

At several points in the trials, carefully crafted (invalid) scoring rubrics were introduced alongside the alternative approach (sometimes in response to students' demands), with very interesting repercussions. Students were generally unable to discern, without assistance, that the preconditions for valid application of assessment rubrics were not being met (i.e., that application of rubrics within their educational context was invalid). Nor could students account for the large variance in grades (determined by aggregate scores) produced by the invalid rubrics (which were not exhibited by the alternative approach). Even so, some students would continue to advocate for the use of assessment rubrics, even in the face of considerable, mounting rationale against the rubrics' contextual validity and educational value.

## **Heightened outcomes**

Some of the advantages that arose from the trialling and the subsequent implementation of the alternative approach include:

- the ability to use the same assessment instrument for students either requiring or preferring personal accommodations;
- tolerance for contextual variability (indeed, students' contextual exploration was greatly enhanced by adopting the alternative instrument);
- attainment of a high level of semi-automation: a bespoke (and readily customisable) guided marking environment, attuned to assessors' workflows, was created by integrating low cost, cloud-based software services;
- rapid uptake in the use of the guided marking environment by new assessors;
- high assessment throughput: complex, contextual, project-based assignments (multi-modal, with sometimes non-linear presentation) were routinely processed within twenty minutes, including the composition of individualised commentary;
- certification (i.e., grade allocations) could be performed reliably within five minutes (on average), with two assessors marking in tandem;
- limited variability in cross-assessor gradings and structured commentary (comment bank) selections, with discrepancies usually confined to a single inter-grade level (typically working to three distinct levels per grade band); and
- very limited contestation of grades and comments by students: usually only one or two cases per semester, with the grounds often linked to external factors (e.g., scholarship conditions or intervention by mentors unfamiliar with the instrument).

## **Accounting for the difference**

The difference in outcomes between the two approaches (assessment rubrics versus the alternative approach) can be accounted for in the following way. Compared to the design requirements of assessment rubrics, the alternative approach:

- relaxes the requirement for orthogonality (no assumption of independent factors);
- respects the constructivist principle (performance need not be predictable);
- adheres to the principle of proximal development (distal statements are not routinely made);
- avoids the need to calibrate factor weights, the weights of ordinal positions, and grade thresholds (constellations are non-mandatory and may vary in their effect);
- removes the assumption of gradated development (there is no requirement for uniform stages within factors, nor co-alignment of ordinal positions across factors); and
- avoids the requirement for complete contextual consistency (greatly expanding the domain of validity).

### **A mathematical account**

Analysing at a more detailed level, the difference between the two approaches can be accounted for in mathematical terms:

- during assessment, the population of an assessment rubric prioritises the act of identification: i.e., matching performance to pre-specified strata, which is an essentially equalising activity;
- the alternative approach prioritises comparison: locating areas of superior and inferior performance, relative to each constellation's ordinal categories;
- instead of looking for equality, the alternative approach shifts to mathematically analytic (i.e., non-equating) enquiry, with assessors looking for the local infimum and supremum (respectively, greatest lower bound and least upper bound) within each constellation;
- with the alternative approach, students' development moves beyond the intensification and attenuation of performance characteristics (of their quantity and quality)—the rubric's implicit model of factor-based capability—replacing this with a dialectical model of conceptual transformation: i.e., the next, proximal performance statement pushes beyond the limits of the current one categorically (i.e., objectively and ontologically); and
- (by contrast) rubrics are statistically categorical or non-parametric—i.e., factor scores and aggregate values are likely to be discrete, non-normal (i.e., essentially arbitrary) probability distributions (Simonoff, 2003)—while with the alternative approach there is no need to impose (or even assume the presence of) a recognised probability distribution per constellation.

### **A sociological account**

Continuing a more detailed analysis of the difference between the two approaches, this time sociologically:

- the emphasis on strata-directed performance equating (per factor) in assessment rubrics helps to perpetuate a restricted world-view comprising pre-established meanings;
- this world-view is inherently normalising, with interpretative deviations and unanticipated (yet still valid) responses being much harder to assimilate;
- if valid deviations are accommodated, the rubric's design framework is unable to adequately justify the accommodations, which undermines the presumed 'objectivity' of the instrument and its surrounding institutional policies;
- the alternative's emphasis on the testing of limits and thresholds (qualitative boundedness per constellation and at a gross level) reflects an approach that is inherently open to novel, creative (and still valid) responses, including those that exceed or depart from existing norms and expectations;

- with the alternative approach, meaning is created dynamically through the assessors' interrogation of the submissions, mediated by testing the connotations of the specification and the assessors' own presuppositions;
- the dialectical sequencing of performance statements in the alternative reflects contemporary understandings of social organisation (e.g., societal forms as variously emergent, dynamic, complex, self-organising, and potentially chaotic); and
- scrutiny of the pragmatic efficacy of the alternative approach's dialectical model leads towards a much richer appreciation of social ontology and its various problematics (i.e., that successful performance involves much more than the intensification/attenuation of desirable/undesirable behaviours).

### **Pedagogical implications**

Given these brief accounts of the differences between the two approaches, a choice made to apply the alternative approach (over assessment rubrics) gives rise to various pedagogical possibilities. For example, the alternative approach encourages use of an interactive pedagogy (primarily student-to-student), which helps the instructor—invoking now a central psychoanalytic concept—to displace the role of the “subject who is supposed to know” (Davis, 1987, p. 750). Within this interactive mode, students can readily share their work, and narrate their emerging practices collaboratively, prior to the submission of assessable work. Similarly, students and assessors are able to explore potentially valid responses prior to the commencement of certification without compromising the assessment process. Content can also be made fairly light (or as laden as required) to help encourage student autonomy (self-directed or cooperative learning) in response to rich assignment specifications.

### **Recommendations for assessment design**

Considering further a broader framework of ‘assessment for learning’ (William, 2011), the alternative approach:

- allows for the under-prescription of assessment requirements, encouraging students' exploration of open possibilities;
- enables, more specifically, vague or ambiguous specification (Thomas, 2008), promoting debate, dialogue, and demarcation: i.e., students taking and defending a position or a particular design approach;
- permits the inclusion in the specification of conflicting or mutually incompatible requirements and constraints (as a means of ensuring students' decision-making);
- prompts designers to prepare pathways for students' creative exploration of disciplinary practices, including an openness and willingness to try interdisciplinary strategies; and

- readily accommodates contextual variation and individual differences and needs within a common, inclusive (and still valid and effective) approach to assessment.

## **Conclusion**

The suggestions offered in this paper are intended to be neither normative nor didactic: instead, they trace the foundations of a general approach towards complex, project-based assessment, with the expectation that evaluation, adaptation, and specific configuration will be required within local settings. Indeed, as the implementation of changes to assessment practices are rarely straightforward, there is a serious risk of negative consequences and potential counter-productivity arising if local, contextual realities are not carefully scrutinised and managed (Richard & Mark, 2010).

In summary, the paper has argued that it is important to acknowledge, articulate, and test the preconditions for valid usage of assessment rubrics, as indeed you would do with any other psychometric instrument. The paper has also defended the claim that the alternative approach presented here has a less restrictive set of preconditions, and therefore maintains (at least theoretically) a wider domain of validity. Trialling and subsequent implementation of instances of the alternative approach has also demonstrated (in context) substantive increases in effectiveness, with significant potential for improving throughput, quality of feedback, and the reliability of grading and certification. In addition, the alternative approach, in comparison with assessment rubrics, lends itself to a more open, responsive pedagogy, which is also scalable, and provides support for student creativity and contextualisation.

This coverage (of validity and effectiveness) also opens onto broader, contemporary topics on assessment and its relationship to learning: topics such as the inter-relationship between assessment approaches and particular value-preferences (Boud & Falchikov, 2007), and the intrinsically interpretative (mediated) nature of performance codification and pre-specification (Sadler, 2014). Being able to use these broader topics to help students and teachers distinguish and differentiate theoretically distinct approaches to assessment becomes crucial in the related task of addressing the issue of ‘pedagogic literacy’ (Price, Handley, Millar, & O’Donovan, 2010). For, it is through a general heightening of pedagogic literacy that students can become important allies in helping institutions to validate (or invalidate) their preferred modes of assessment—particularly through students’ ability to generate rapid shifts in scale and new orders of magnitude.

Of course, this paper has raised—and diplomatically avoided—the question of how this entire situation has arisen (of the prevalence of assessment rubrics), but to conclude, the answer lies principally within the sociological domain, with significant conceptual and pragmatic transformations required to break the appearance of its naturalisation. The attempt made in this paper to commence this process has relied, as we’ve seen, on multiple academic disciplines, and it can be expected that further interdisciplinary interventions (or the adoption of a general, transdisciplinary approach) will be required

to make further progress. This task would also benefit from the establishment of multiple, cross-institutional collaborations to help promote a better understanding of educational contextualisation and its related effects (Bernstein, 2000).

## References

- Allen, D., & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE: Life Sciences Education*, 5(3), 197-203.
- Andrade, H. (2007). Self-assessment through rubrics. *Educational Leadership*, 65(4), 60-63.
- Baryla, E., Shelley, G., & Trainor, W. (2012). Transforming rubrics using factor analysis. *Practical Assessment, Research & Evaluation*, 17(4), 1-7.
- Bernstein, B. B. (2000). *Pedagogy, symbolic control, and identity: Theory, research, critique* (rev. ed.). Lanham, MD: Rowman & Littlefield.
- Boud, D., & Falchikov, N. (Eds.). (2007). *Rethinking assessment in higher education: Learning for the longer term*. London, UK: Routledge.
- Connelly, M., & Wolf, K. (2007). The art of rubrics: Painting by numbers? *Teaching Artist Journal*, 5(4), 279-288.
- Davis, R. C. (1987). Pedagogy, Lacan, and the Freudian subject. *College English*, 49(7), 749-755.
- Goodrich, H. (1996). Understanding rubrics. *Educational Leadership*, 54(4), 14-17.
- Goodrich Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-18.
- Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the past into the future. *Journal of Applied Linguistics*, 1(1), 49-72.
- Martin, J. R. (1992). *English text: System and structure*. Amsterdam, Netherlands: John Benjamins.
- Martin, J. R., & Rose, D. (2008). *Genre relations: Mapping culture*. London, UK: Equinox.
- Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment, Research & Evaluation*, 7(3).
- Oxford, R. L. (1997). Constructivism: Shape-shifting, substance, and teacher education applications. *Peabody Journal of Education*, 72(1), 35-66.
- Poehner, M. E. (2012). The zone of proximal development and the genesis of self-assessment. *The Modern Language Journal*, 96(4), 610-622.
- Popham, W. J. (1997). What's wrong—and what's right—with rubrics. *Educational Leadership*, 55(2), 72-75.
- Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment & Evaluation in Higher Education*, 35(3), 277-289.

- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448.
- Richard, B., & Mark, G. (2010). Is the feedback in higher education assessment worth the paper it is written on? Teachers' reflections on their practices. *Teaching in Higher Education*, 15(2), 187-198.
- Sadler, D. R. (2014). The futility of attempting to codify academic achievement standards. *Higher Education*, 67(3), 273-288.
- Simonoff, J. S. (2003). *Analyzing categorical data*. New York, NY: Springer.
- Stevenson, A. (Ed.). (2010). *The Oxford dictionary of English*. (3rd ed.). Oxford, UK: Oxford University Press.
- Thomas, K. (2008). Ambiguity as a hallmark of pedagogical exchanges between art teachers and students in the making of creative artworks. *Australian Art Education*, 31(2), 4-21.
- 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).
- William, D. (2011). What is assessment for learning? *Studies In Educational Evaluation*, 37(1), 3-14.