

Improving the Quality of Assessment Grading Tools in Master of Education Courses: A Comparative Case Study in the Scholarship of Teaching and Learning

**Michael Christie¹, Peter Grainger², Robert Dahlgren³, Kairen Call⁴,
Deborah Heck⁵, Susan Simon⁶**

Abstract: This study compares the use and efficacy of assessment grading tools within postgraduate education courses in a regional Australian university and a regional university in the US. Specifically, we investigate how the quality of postgraduate education courses can be improved through the use of assessment rubrics or criterion referenced assessment sheets (CRA sheets). The researchers used a critical review of rubrics from Master of Education courses, interviews and a modified form of the Delphi method to investigate how one can assure the quality of assessment grading tools and their effects on student motivation and learning. The research resulted in the development of a checklist, in the form of a set of questions, that lecturers should ask themselves before writing rubrics or CRA sheets. The paper demonstrates how assessment grading tools might be researched, developed, applied and constantly improved in order to advance the Scholarship of Teaching and Learning.

Key words: Criterion referenced assessment; grading tools; criteria sheets; Delphi technique/model

Introduction

We need to begin by defining our terms and clarifying the features of criterion referenced assessment (CRA). In Australia and the US the tool used in CRA is commonly called an assessment criteria sheet or rubric. An online search of 20 teaching and learning centre websites in both US and Australian universities (27 April 2015) revealed that both terms were used interchangeably. We will do the same in this article. A rubric is a tool for interpreting and judging students' work against set criteria and standards. The rubric is often presented as a matrix or a grid but there are other, arguably better models, for presenting a rubric. Grainger and Weir (2015) evaluated two styles of criteria sheets: the traditional matrix style criteria sheet and the Continua model of a Guide to Making Judgements (GTMJ). More research in this area is desirable. In principle the purpose of a rubric is to make explicit the range of assessment criteria and expected performance standards for a task or performance. The assessor evaluates and identifies the standard of what a student has submitted against each of the individual assessment criteria and provides an overall judgment for the task or performance as a whole. Another term that we need to define, since it underpins the whole

¹ School of Education, University of the Sunshine Coast, 90 Sippy Downs Dr., Queensland, 4556, Australia.

² School of Education, University of the Sunshine Coast, Queensland, 4558, Australia.

³ Fredonia College of Education, State University of New York, 280 Central, Fredonia, NY 14063, United States, New York.

⁴ School of Education, University of the Sunshine Coast, 90 Sippy Downs Dr., Queensland, 4556, Australia.

⁵ School of Education, University of the Sunshine Coast, 90 Sippy Downs Dr., Queensland, 4556, Australia.

⁶ School of Education, University of the Sunshine Coast, 90 Sippy Downs Dr., Queensland, 4556, Australia.

case study, is quality. We have decided, for the purposes of this study, to define *quality* by means of a hybrid of two common definitions. For us *quality* is best characterised as fitness for purpose and constant improvement.

In a series of articles and keynote addresses, that span almost two decades, Sadler (1987, 2005, 2007, 2009, 2010, 2013) argued that educational institutions are becoming more committed to using criterion referenced assessment in order to promote effective student learning. He provided convincing evidence in the articles that focused specifically on higher education (Sadler, 2005, 2009) on the connection between good rubrics and good learning. This paper provides a specific, comparative case that helps substantiate the assumption that CRA and well written rubrics will increase the quality of learning. Well composed rubrics not only help the student but also force the teacher to be more exact in the formulation of learning tasks. They also simplify moderation processes because moderators use a common set of criteria to judge a piece of work. Rubrics are efficacious in that they do good during their creation as well as their application. The best way to develop and use them is collaboratively. Involving one's peers as well as one's students in the construction and application of rubrics is a cornerstone of CRA. Jonsson (2014) identified that rubrics made assessment tasks more transparent for students and provided them with the tools to unlock secret by involving them in the assessment process. Rubrics provide students with greater ownership and understanding of the rubric providing the option to undertaken self-assessment. This is something we have endeavoured to do in our case study. The fact that the fourth author was a student in the courses that make up the Australian part of our case study indicates our commitment to involving academic staff and students in the process.

Our study was conducted as part of an international peer review project carried out during 2014-2015 by a team of educational researchers from a regional Australian University and their colleagues from a similar sized, regional tertiary institution in the United States (US). The project used the acronym PEER which stands for Postgraduate Evaluation of Educational Research. Although funding was minimal the aim of the project was ambitious, namely to develop a transferable, online, blended learning model of peer review for research-related Masters of Education degree courses. The model was designed to improve the quality of students' verbal and written reports and save universities time and money. The project involved six lecturers and seventy Master of Education students from both institutions. We divided the project into three sub projects, namely a project focusing on online exchange and review of presentations, improving professional peer review in leadership courses and a project where colleagues from the two universities carried out a case study to improve the efficacy of rubrics, particularly in project-based MEd courses. It is this third sub-project that is reported on here.

Comparative Policies Regarding CRA and Rubrics in Australia and the US

In Australia university lecturers are finding that, whether they like it or not, criterion referenced assessment and the associated use of rubrics, is being directly regulated from above. Government in Australia subsidizes universities and, understandably, creates agencies to ensure that taxpayer money is being spent on a quality product. The Bradley Report (Department of Education Employment and Workplace Relations, 2008) resulted in the Australian Federal Government setting up a new agency for regulating universities called the Tertiary Education Quality and Standards Authority, or TEQSA. A key focus for TEQSA is the development of a set of threshold standards for every level of program offered at any Australian university. These are outlined in the Higher Education Standards Framework (Department of Industry Innovation Science Research and Tertiary Education, 2011).

These reforms include an opportunity for universities to investigate alternative assessment frameworks that can accommodate TEQSA's new standards-based assessment mandates. According to item 5.5 of the TEQSA framework (Department of Industry Innovation Science Research and Tertiary Education, 2011, p. 16) there is a requirement to benchmark standards against similar accredited courses of study offered by other higher education providers. In order to carry out this type of institutional benchmarking universities need a common understanding of assessment principles (Boud & Associates, 2010). This includes the use of rubrics. Top down reforms have a knock-on effect. To comply with TEQSA universities, in their turn, mandate the use of course outlines that include assessment criteria for course tasks and tests. Most lecturers feel obliged to develop rubrics that show how students will be judged according to the criteria. The most common rubric they use is the Matrix style shown in figure 1 below, although it is possible to use variations to this model, for example, the 'guide to making judgments' or continua model (see appendix A).

	Standard A	Standard B	Standard C	Standard D	Standard E
Criterion 1	Standard descriptor	Standard descriptor	Standard descriptor	Standard descriptor	Standard descriptor
Criterion 2	Standard descriptor	Standard descriptor	Standard descriptor	Standard descriptor	Standard descriptor

Figure 1: Matrix model. Source: Authors

In Australian universities the standards typically refer to High Distinction, Distinction, Credit, Pass, and Fail. Writing the standard descriptors is a challenging task for lecturers who may not be assessment experts. If a criterion for an essay is, for example, that it displays a 'logical argument' the lecturer might resort to using a set of adjectives, such as an 'excellent, very good, good, passable and incoherent' to explain the standard, which leaves the student wondering how the assessor will distinguish between these terms. The use of rubrics in Australia and the US gained significant support towards the end of last century, particularly in schools, but as Popham (1997) asserted, in a provocative article in *Educational Leadership*, '... the vast majority of rubrics are instructionally fraudulent' (p.73). Popham was talking, in the main, about commercially produced rubrics for schools, but many of the points he made in his article remain valid today, particularly in universities.

The United States, in contrast to Australia, does not have a National Authority for regulating quality in higher education institutions. This work is left to accrediting bodies for institutions such as the Accrediting Council for Independent Colleges and Schools (ACICS) as well as for disciplines, for instance, ABET which stands for the Accreditation Board of Engineering and Technology. The US Department of Education takes a more federalist approach toward governing public institutions of higher education. It offers a modicum of support but leaves administrative matters in the hands of the respective state governments. In the discipline of Education, despite recent efforts at standardization, this approach has led to differences in the way states enforce standards for initial teacher education programs and Master of Education courses.

Our project partners at SUNY Fredonia's College of Education teach in pre and in service teacher education courses. Their courses exemplify how differences, between a national versus state accreditation system, can affect assessment and assessment rubrics in *Journal of the Scholarship of Teaching and Learning*, Vol. 15, No. 5, October, 2015.
Josotl.Indiana.edu

Australia and the US. All initial teacher education programs in Australia not only need to meet TEQSA standards, but in addition devise tasks that enable their students to prove that they have meet the seven standards mandated by the Australian Institute for Teaching and School Leadership (AITSL). The tasks are rarely multiple choice and short answer tests, but they must be published in course outlines that clearly state the criteria by which they will be assessed. These can be audited and universities can lose the right to graduate teachers if they requirements are not met. Graduates from accredited courses have the right to register as teachers via an administrative process in each state.

In New York State the pre-service teachers are required to take a number of New York State Education Department (NYDED) tests, after graduation, in order to gain teacher registration. The tests are composed of multiple choice and short answer questions and are designed to assure the quality of a prospective teacher by checking their knowledge and skills in pedagogy, academic literacy, subject speciality and diversity awareness, among other things. The tests are professionally produced and rubrics explaining how they are marked are available online. For example, in the Academic Skills Literacy Test, the marking rubric for the criterion connected to argumentative writing skills is as follows:

Score Point	Score Point Description
4	The "4" response demonstrates a strong command of argumentative writing skills.
3	The "3" response demonstrates a satisfactory command of argumentative writing skills.
2	The "2" response demonstrates limited argumentative writing skills.
1	The "1" response demonstrates a lack of argumentative writing skills.
U	The response is unscorable because it is unrelated to the assigned topic or off-task, unreadable, written in a language other than English or contains an insufficient amount of original work to score.
B	No response.

Figure 2: Extract from rubric for ALST. Source: NY State Education Department.

For this particular criterion the descriptors are not so different from our earlier example, and again, one would like to know in what way exactly does a student demonstrate ‘a strong command of argumentative writing skills’. Once registered, a new teacher must, within a five-year period, obtain a Master’s degree in order to continue their certification beyond the initial level. Given the mix of private and state higher education institutions, capstone assignments for the Masters of Education can vary. Within the State University of New York (SUNY) system, which is made up of 64 institutions, a standard thesis acts as a capstone assignment for advanced teacher preparation. Each institution has the latitude to choose the sequence of courses and assignments that faculty thinks best supports the candidates in the writing of their theses. The most common is a three-course sequence involving an introduction to educational research, a course during which students develop thesis proposals and a final capstone course in which candidates collect and analyse the data from their projects and complete the written requirements for the thesis. The lecturers for each course can decide to produce rubrics or not. In our sub project three of the US team had done so and one had not. The style and quality of the rubrics also varied which we discuss below.

The Problem and How to Deal With It

The current emphasis on standards creates new challenges for tertiary educators. They and their institutions need to rethink and renew the tools they use to assess learning if they are to

be a help to learning rather than a hindrance. The problem that our paper addresses is that Popham (1997) diatribe against potentially educationally fraudulent rubrics can be levelled at those being devised by lecturers in undergraduate and postgraduate courses in Australian and US universities. There is no deliberate intention to 'defraud', but in their haste, lecturers are prone to mistake the performance test of a skill for the skill itself and write rubrics that specifically address the criteria relevant to the task or test, rather than the skill. The criteria and the standard descriptors must be general enough that they could be used with another performance test of that skill. On the other hand they should not be so general, as the descriptors of argumentative writing in the NYSED tests are, that there is no clear indication of what one must do 'to demonstrate a strong command of argumentative writing skills'.

Australian and US academics need support in developing the expertise required to take on new and demanding assessment responsibilities intended to assist benchmarking and quality assurance of standards in tertiary education (Boud & Associates, 2010). Our case study helps develop a common language for describing and interpreting assessment criteria and standards, and presents a checklist that lecturers can ask themselves before designing, developing and improving their rubrics. The literature shows that there is a causal connection between the use of well constructed rubrics and increased understanding and learning on the part of students. Panadero and Jonsson (2013), after analysing 21 studies on rubrics, found that rubrics '...have the potential to influence students learning positively' and that 'there are several different ways for the use of rubrics to mediate improved performance and self-regulation' (p.129). In another meta review of rubric use in higher education, Reddy and Andrade (2010) made the important point that students and their lecturers have different perceptions of the purpose of rubrics. The former saw them as assisting learning and achievement whereas their teachers were much more focussed on the role of rubrics in 'quickly, objectively and accurately assigning grades' (p.5). In the USA, at least, their review of the literature reveals a reluctance on the part of college and university teachers to use rubrics. Reddy and Andrade (2010) suggest that lecturers might be more receptive if 'they understand that rubrics can be used to enhance teaching and learning as well as to evaluate it' (p.439). In other words, rubrics need to be seen as formative as well summative in their purpose (Clarke, 2005; Clarke, Timperley, & Hattie, 2004; Glaser, 2014; Glasson, 2009). In our case study we use qualitative research methods to create a checklist of questions that lecturers can ask themselves before writing rubrics or CRA sheets. The paper demonstrates how assessment grading tools might be researched, developed, applied and constantly improved in order to advance the Scholarship of Teaching and Learning.

Methodology

In our case study we combined a search of the literature with three in-depth interviews and two rounds of a modified Delphi Method. The interviews focused on whether good rubrics can motivate and assist the learning of postgraduate students, many of whom are professionals returning to study a MEd course. The interviewees in this study consisted of an Australian expert in assessment, a US lecturer in a MEd course and an Australian student who had recently completed a MEd by coursework. Because of logistics the interviewees responded to the questions via email. We used an analysis of the interview responses to develop a number of themes and pertinent questions connected with the development and quality assurance of rubrics.

The Delphi method has been used extensively in participatory action research although its origins date back to the cold war when it was used extensively as a forecasting mechanism by the Rand Project (Brown, 1968). We modified the Delphi method in that the first set of guiding questions were produced by the authors, who after an analysis of the *Journal of the Scholarship of Teaching and Learning*, Vol. 15, No. 5, October, 2015.
Josotl.Indiana.edu

interviews and the survey responses, wrote down a set of questions. This first provided a total of 41 questions. These responses were reduced to 20 guiding questions and these were sent out for a second round and the individual respondents were asked to look at them and come up with their best five questions. Their responses (30) were filtered using the same principles of overlap to produce a final checklist of the best ten questions that a lecturer could ask before writing a rubric. To conclude the process the set of 10 questions were sent out to three experts who were chosen because they had published a number of articles on assessment and in the case of two, edited a book on the subject. Some modifications were made on the basis of their response.

Our modified Delphi was designed as a useful methodological adaptation for university academics interested in developing their own Scholarship of Teaching and Learning (SOTL). Although the sorting method has some resemblance to the constant comparison method in grounded theory it differs in that the goal is to reach a consensus on a predetermined issue rather than to build theory. In our Delphi exercise we looked for conceptual similarities, refined categories and looked for patterns (Tesch, 1990) which are all part of a grounded theory approach but our research was applied rather than theoretical.

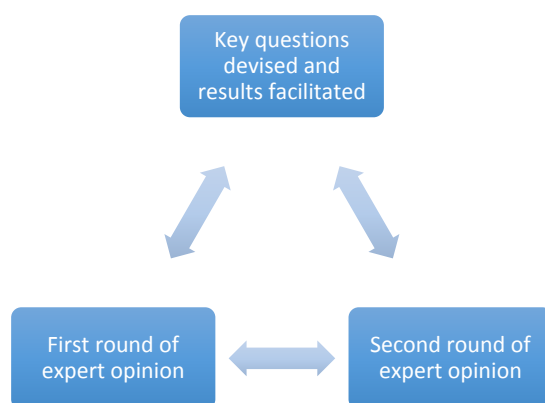


Figure 3. Adapted Model of Delphi Method. Source: Authors

Data Collection and Analysis

Assessment can foster and drive student learning. However, in higher education where there is so much emphasis on grading via written tests and exams the quality of assessment can lead to either surface or deep approaches to learning (Biggs, 2001; Hounsell, 2005). Because higher education is increasingly a form of professional training for teachers, nurses, doctors, scientists, engineers and so many other professions, assuring the quality of that professional preparation is essential. As a result, there has been a renewed focus on improving assessment practice in tertiary education because of its powerful impact on the quality of learning and eventually the quality of the people inducted into different professions (Biggs, 2001; Boud & Associates, 2010). Responses from our interviewees stressed the efficacy of quality rubrics to encourage a deep approach to learning and a sufficient understanding to apply knowledge and skills in a variety of settings.

The three interviewees, represented here by the initials AS (Australian Student), AE (Australian Expert) and AL (American Lecturer), were largely in agreement on a number of points. Their responses, encapsulated in the body of emails and attachments resonated with findings in the literature. AS and AE emphasized the importance of using high quality rubrics in conjunction with assessable tasks. AS said that for students, assessment criteria are integral to their understanding of tasks and success in undertaking them. This is a perspective that

deserves more research in the literature. AS had just completed the required courses for a Masters of Education and reported that fellow students spoke highly of good quality rubrics because of the transparency they provided in terms of the task requirements. The key here is the quality of the rubrics, a point that was underscored in AE's response. Poor quality assessment sheets or rubrics that do not fit their proclaimed purpose can be misleading and confusing rather than motivating.

According to AS the quality and use of rubrics in the courses, including those that are the focus of our case study, varied. In comparing rubrics all three respondents raised a number of key issues that throw light on how CRA and rubrics can help or hinder learning. AS criticised the lack of consistency in formatting, interpretation and approach taken by lecturers but made the observation that these differences meant that engaged students discussed and critically reflected on the strengths and weaknesses of the criteria sheets. The result of such peer review was positive according to AS, but clearly the person who wrote the rubric should have also been involved if we are to accept the findings of Eshun and Osei-Poku (2013), whose study involving 108 university students revealed that students need training in the use of rubrics. In fairness AS did say that certain lecturers discussed the rubric together with the students and made adjustments to it where there were obvious weaknesses.

In AE's response a Continua model of a guide for making judgments or the GTMJ model was presented (see Appendix A). According to AE this type of rubric was becoming more common in the program that is the focus of our case study. The matrix rubrics experienced by AS used *High Distinction* (HD) through to a *Fail* grade in the header for the standards, but some other lecturers used terms such as *Exceptional* through to *Unsatisfactory*. In the response from AL an example of a rubric for an annotated bibliography task was cited. This used *A Excellent*, *A- Great bibliography*, *B+ Very good bibliography*, *B Good bibliography*, *B- Fair bibliography*, *C Poor bibliography*, and *E Unable to complete assignment*. To compound the problem, according to all three informants, the actual marks that matched the letters were rarely given on the criteria assessment sheet. In most cases, students had to find out what the letters meant in terms of marks from another source.

In the rubrics cited by AS most lecturers provided descriptors for all grade levels from a High Distinction (HD) through to a Fail grade. However, a number of criteria sheets neglected to offer a descriptor below a Pass level, which meant failing students were left outside of the framework. Standard descriptors are a significant reference point for students, according to AS, both during the task development and feedback phases and as such, clarification of the messages within them is essential. According to AE and AL the standard descriptor needs to explain what has to be done using a verb that incorporates the higher level of learning achieved. AS pointed out that it was unhelpful to have a criterion for a task such as 'understands x' and then just add a descriptor under, for example the *HD* column which says 'demonstrates *Excellent* understanding of x'. This is compounded when other adjectives such as *Very Good*, *Good*, *Satisfactory* and *Unsatisfactory* are used in the other grade columns with no indication as to how excellent or satisfactory understanding is actually demonstrated. As AE pointed out, one needs to integrate a taxonomy, such as Bloom, Engelhart, Furst, Hill, and Krathwohl (1956) so that the quality of understanding can be judged by whether or not one has done certain, specified things that demonstrate for example if the student is capable only of declarative knowledge as opposed to being able to contrast, compare and evaluate aspects of that knowledge.

In the studies AS undertook, some criteria sheet formats offered descriptors at only the highest and lowest standards. AS argued that while they contained less detail, the quality of information was sufficient to clearly guide the learning process. According to AS this format placed 'greater emphasis on the criteria themselves rather than the range of standard descriptors, providing scope for differences in approach, creativity and personal style'. AS *Journal of the Scholarship of Teaching and Learning*, Vol. 15, No. 5, October, 2015.
Josotl.Indiana.edu

added the proviso that ‘this format may become problematic when a student attempts to determine why they received a certain grade, and as such its success relies heavily on the assessor providing detailed written feedback’. Both AS and AE mentioned the Masters level skills identified by the Australian Qualifications Framework (AQF) (Australian Qualifications Framework Council, 2011) and raised the question of how the standards descriptors support the broader AQF level descriptors for Master of Education students? AS pointed out the dilemma of finding a balance between highly specific rubrics that provide detailed standard descriptors for all levels (matrix model) or the type mentioned above that only gives the descriptors for the top and bottom standards. According to AS the matrix model ‘gives clear indicators for success during the task production phase and a comprehensive checklist within the feedback phase’. AS cautioned that this model ‘can divert attention away from learning and towards deconstructing the complexities of the criteria involved’. It can also ‘lead the student to believe that the assessor has a specific product in mind’.

Both AE and AL said that they engaged students in a discussion about the rubrics they wrote for their specific course tasks. This was important for students, according to AS who said that interpretation of criteria was a regular feature of discussion within classes throughout the program. All three agreed that when discussion about criteria forms part of the learning, from the start of the course, misunderstandings are reduced. The interviewees all mentioned the problematic nature of inherited rubrics, where the assessor has taken over someone else’s course and its assessment rubrics. In that case both assessor and student need to interpret the criteria and standard descriptors. In the cases AS experienced, assessors worked with students to create a shared definition and understanding, aligning the course learning objectives to the assessment criteria. This highlights the need for criteria sheets to be regularly peer reviewed at the faculty level, in order to ensure clarity beyond the author of the criteria sheet.

The interview responses from AE and AS, both of whom were involved with the MED program that is the focus of our study, stressed the importance of face-to-face feedback to students. They noted that a common practice in the written feedback was to fill out a form composed of the rubric itself with the descriptors within specific standards highlighted and then give a brief, general comment in a lined space beneath the rubric. AS said, that from the student perspective, this offered a precise understanding of where a student sits within the university grading scale but if a descriptor contains several components it can be difficult for a student to determine their level of success. In order to navigate this, and offer students more specific feedback, some assessors highlighted parts of descriptors across different standards. This served to demonstrate that the lines between standard descriptors are not solid, but rather work as a continuum. AS would have preferred a consensus from lecturers in the use of criteria sheets in the feedback phase. A common approach would enable students to engage with the feedback more effectively, rather than seeking clarification from individual lecturers.

In our modified Delphi the forty one responses from the first round covered issues and questions similar to those raised in the interviews. Themes were identified within the 41 original responses which enabled us to reduce them to a set of 20 guiding questions. Each expert was then asked to examine the 20 guiding questions and individually produce a set of the most significant five. The resulting list of 30 questions, which naturally contained considerable overlap was then reduced to the following questions which can be used by academics to develop and evaluate the quality of rubrics or criteria sheets. They are:

1. Does the rubric have criteria that are clear/unambiguous?
2. Do the criteria explain what must be done and demonstrated?
3. Are the criteria knowledge based and skills based at a Masters level standard?

Journal of the Scholarship of Teaching and Learning, Vol. 15, No. 5, October, 2015.

Josotl.Indiana.edu

4. Does the criteria sheet have standards identified (i.e., HD, D, C, P, F)?
5. Are the standards' descriptors explicit, devoid of subjective words, and positively worded in terms of what students must do?
6. Are there gradations of quality that differentiate the standards clearly, for example, according to a taxonomy of learning such as Bloom's taxonomy?
7. Is the layout of the criteria sheet clear, not too crowded, uncluttered, nested?
8. Does the task provide opportunities for the students to demonstrate that they have achieved its intended outcomes, graduate attributes and skills according to specific criteria?
9. Does the rubric reflect what students have studied for the task and enable them to demonstrate that they have met its criteria and standards?
10. Does the rubric reflect course outlines as well as graduate attributes and skills?

Results and Discussion

The project revealed significant differences both within and between Australian and US practices when it comes to the use of rubrics in Master of Education courses. The lack of standardization, internally and externally within Master of Education courses at both institutions, is reflected in the variety of grading tools used to mark student work. In our case study, the US lecturers who took Master of Education courses, all used different assessment schedules whereas their Australian counterparts uniformly adhered to CRA and most used a matrix model criteria sheet. One used the continua model of a Guide to Making Judgments mentioned above and exemplified in Appendix A.

We argue that Master of Education courses can be improved, both in Australia and the USA, via a shared understanding of assessment principles and a reform of existing assessment practices, including the instruments used to grade student work. The key is that the tools used to evaluate student learning are truly criterion referenced and standards based, where 'standards are set above the norm with a high achievement focus' (Gittens, 2007, p. 2). Shifts to a standards-based curriculum framework in teaching and learning are in keeping with national and international efforts to standardize and assure research quality. Australia's higher education accrediting agency, TEQSA, will place increasing pressure on lecturers, their departments and their institutions to conform to standardized assessment regimes. Grading tools are a key to quality assurance but our research has highlighted that their design and efficacy for judging student work often varies within and across tertiary education contexts.

In the US, at least from evidence in our case study, there is much more scope for individuality when it comes to writing rubrics. AL conceded that there was 'a good deal of latitude for individual instructors in terms of how they organize their courses' including the writing of rubrics. Fredonia's College of Education (COE), on the advice of faculty working parties, has compiled a handbook on graduate research in education that standardizes the thesis components and submission guidelines. However the development of rubrics, and appraisal of their validity, remains with the individual lecturers. In those instances where rubrics are not used the lecturers explain that they use their professional judgment to allot grades. The use of professional judgement as a quality assurance measurement in the US is partially supported in research by (Banta & Palomba, 2014; Connolly, Klenowski, & Wyatt-Smith, 2012; Klenowski & Adie, 2009; Race, 2006; Readman & Allen, 2013; Sadler, 2013). They indicate that academics who are experienced assessors possess tacit knowledge of what quality in student work looks like. Sadler demonstrated that competent appraisers can consistently identify quality when they see it. This tacit knowledge has been shown to enable assessors to make accurate interpretations of sometimes vague descriptions of student

behaviour in order to discriminate between standards or levels of achievement (Grainger, Purnell, & Zipf, 2008). In some respects professional judgment can act as a fail-safe mechanism to help ensure that experienced lecturers, who inherit defective criteria sheets, can make adjustments so that there is no compromise of assessment integrity and reliability in judging student work. Naturally such lecturers need to rewrite the rubric as soon as possible.

In Australia the matrix style grading tool is commonly used but we have argued throughout this paper that its value depends on the quality of its criteria, standards and standard descriptors. Not all academics understand the rigor needed with criteria and standards based assessment, and it takes some years to get to know how to consistently align evidence of quality with relevant achievement standards. For assessors who are unclear about learning quality, vague assessment rubrics can mitigate against objective judgment of performance and undermine consistency of teacher judgments. Grading tool deficiencies represent a major challenge to what Sadler (2010) refers to as 'grade integrity'. Completely objective judgements of performance become impossible. That is why moderation of grades is necessary. However, it is desirable to aim for the optimum level of clarity in the standards descriptors in grading tools in order to enhance the moderation process.



Criteria sheets or rubrics are meant to enable assessors to evaluate the quality of student work as well as guide student learning by making explicit the evidence needed to demonstrate the requirements of the assessment task. These requirements are typically defined in the standards descriptors. Because standards descriptors have more than one purpose and audience, they are not easy to construct to adequately differentiate between levels of achievement. This can result in descriptions of standards that are vague, unclear, indicative only and open to interpretation. Too often it is assumed that the student will be familiar with and understand the language used in the descriptors. Sadler (1987, 2009) argues that standards descriptors must be precise to allow for unambiguous determinations and they must consist of statements that accurately describe the properties which characterise a learning behaviour at its designated level of quality.

We have shown that ambiguous descriptors are problematic for both marker and student, because the required behaviours are vague. The implication for marking is that assessors may be encouraged to ignore the standards descriptors and evaluate student work based on their own criteria, which brings into question the integrity of the final judgement. Evidence of this is reported by Klenowski and Adie (2009). Another major discussion point, raised in both the interviews and Delphi responses, is the issue of alignment. Firstly, alignment of the task and the criteria sheet with the relevant course outline, and then alignment with the graduate attributes and institutional and national requirements.

Assessment is the making of judgments about how students' work aligns with appropriate standards. It serves a number of purposes, including certification, but in terms of learning it should also help students to identify and engage in quality learning (Boud & Associates, 2010). If students are not able to do this as a result of poor assessment practices, the educational purpose of assessment is lost. Rubrics are designed to help assessors make judgments about quality, and justify that quality by using appropriate standards descriptors. They are also an excellent mechanism for giving detailed feedback to students. Boud and Associates (2010) point out that we need specific and detailed information in order to show students what they have done well or not, and how their work could be better. To design, develop and improve on rubrics one needs to ask the right questions. The set of questions that we offer as the result of our study were part of a collegial, international exercise in the scholarship of teaching and learning. Our intention is to make use of the questions to improve on our own rubrics and instigate another cycle of research to see to what extent our students perceive that the revised rubrics help them in their learning. If others follow our example,

then the scholarship of teaching and learning in this area can be shared and deepened in both Australia and the US.

Appendix 1. Example of a Continua Model of a GTMJ. Source: Authors

Knowledge and understanding	Ways of working/Skills	
Knowledge and understanding of curriculum development	Academic literacies referring to referencing English expression, use of literature, spelling, grammar, punctuation	
 <ul style="list-style-type: none"> Justifies a variety of aspects of the curriculum in detail. Discusses a variety of different aspects of the curriculum in detail Identifies the key or fundamental aspects of the curriculum Writes brief, fragmented, superficial facts about the curriculum 	 <ul style="list-style-type: none"> Makes links between paragraphs to ensure continuity. Uses sources to enhance arguments. Writes consistently accurate references. Writes with isolated technical errors. Critically analyses sources by comparing and contrasting the views of many different authors to support arguments. Writes with minor technical errors. Writes an accurate and formal introduction and conclusion explaining the discussion framework. Logical sequence of content. Cites a variety of different sources to justify statements including the most recognised experts. Writes using recognizable APA style, following the key conventions consistently. Makes a frequent variety of technical errors that don't impede understanding. Recognisable formal introduction and conclusion. Cites key sources . Writes with many different types of key technical errors that distort meaning. Cites unrecognised sources. Consistently makes statements that are not supported by sources. 	<div>HD</div> <div>D</div> <div>C</div> <div>P</div> <div>F</div>

References

- Australian Qualifications Framework Council. (2011). Australian Qualifications Framework Retrieved January 12, 2015, from http://www.aqf.edu.au/Portals/0/Documents/Handbook/AustQuals%20FrmwrkFirstEitionJul2011_FINAL.pdf
- Banta, T. W., & Palomba, C. A. (2014). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education* (2nd Ed.). Hoboken: Wiley.
- Biggs, J. (2001). The reflective institution: Assuring and enhancing the quality of teaching and learning. *Higher Education*, 41(3), 221-238. doi: 10.1023/A:1004181331049
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive Domain*. New York: David McKay Company.
- Boud, D., & Associates. (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Sydney: Australian Learning and Teaching Council. Retrieved March 21, from http://www.uts.edu.au/sites/default/files/Assessment-2020_propositions_final.pdf.
- Brown, B. B. (1968). *Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts*. Santa Monica, California: The RAND Corporation. Retrieved January 12, 2015, from <https://www.rand.org/content/dam/rand/pubs/papers/2006/P3925.pdf>.
- Clarke, S. (2005). *Formative assessment in the secondary classroom*. London: Hodder Murraby.
- Clarke, S., Timperley, H., & Hattie, J. (2004). *Unlocking formative assessment: Practical strategies for enhancing students' learning in the primary and intermediate classroom*. Auckland: Hodder Education.
- Connolly, S., Klenowski, V., & Wyatt-Smith, C. M. (2012). Moderation and Consistency of Teacher Judgement: Teachers' Views. *British Educational Research Journal*, 38 (4), 593-614. Retrieved February 16, 2015, from <http://dx.doi.org/10.1080/01411926.2011.569006>
- Department of Education Employment and Workplace Relations. (2008). *Review of Australian Higher Education: Final Report (Bradley Review)*. (978-0-642-77805-5). Canberra: Australian Government Retrieved November 29, 2014, from <http://www.mq.edu.au/pubstatic/public/download.jsp?id=111997>.
- Department of Industry Innovation Science Research and Tertiary Education. (2011). *Higher education standards framework* (F2013L00194). Canberra: Australian Government Retrieved November 29, 2014, from <http://www.comlaw.gov.au/Details/F2013C00169/Download>
- Eshun, E. F., & Osei-Poku, P. (2013). Design Students Perspectives on Assessment Rubric in Studio-Based Learning. *Journal of University Teaching and Learning Practice*, 10 (1).
- Journal of the Scholarship of Teaching and Learning*, Vol. 15, No. 5, October, 2015.
Josotl.Indiana.edu

Retrieved, January 12, 2015, from
<http://ezproxy.usc.edu.au:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1005281&site=eds-live>

Gittens, W. (2007). Shifting Discourse in College Teaching. *International Journal for the Scholarship of Teaching and Learning*, 1(1), Article 17. Retrieved February 16, 2015, from
<http://digitalcommons.georgiasouthern.edu/ij-sotl/vol1/iss1/17>

Glaser, R. E. (2014). Design and Assessment of an Assignment-Based Curriculum to Teach Scientific Writing and Scientific Peer Review. *Journal of Learning Design*, 7 (2), 85-104. Retrieved February 16, 2015, from
<http://ezproxy.usc.edu.au:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1040438&site=eds-live>

Glasson, T. (2009). Improving student achievement: A practical guide to assessment for learning. Carlton South Vic.: Curriculum Corp.

Grainger, P., Purnell, K., & Zipf, R. (2008). Judging quality through substantive conversations between markers. *Assessment & Evaluation in Higher Education*, 33 (2), 133-142. doi: 10.1080/02602930601125681

Grainger, P., & Weir, K. (2015). An alternative grading tool for enhancing assessment practice and quality assurance in higher education. *Innovations in Education and Teaching International*. doi: 10.1080/14703297.2015.1022200

Hounsell, D. (2005). Contrasting conceptions of essay-writing. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The Experience of Learning: Implications for teaching and studying in higher education* (3rd (Internet) ed., pp. 106-125). Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment. Retrieved January 27, 2015, from
http://www.docs.hss.ed.ac.uk/iad/Learning_teaching/Academic_teaching/Resources/Experience_of_learning/EoLChapter7.pdf.

Jonsson, A. (2014). Rubrics as a Way of Providing Transparency in Assessment. *Assessment & Evaluation in Higher Education*, 39 (7), 840-852.

Klenowski, V., & Adie, L. (2009). Moderation as judgement practice: reconciling system level accountability and local level practice. *Curriculum Perspectives*, 29 (1), 10-28. Retrieved November 29, 2014, from <http://eprints.qut.edu.au/26164/>

Panadero, E., & Jonsson, A. (2013). Review: The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144. doi: 10.1016/j.edurev.2013.01.002

Popham, W. J. (1997). What's wrong--and what's right--with rubrics. *Educational Leadership*, 55 (2), 72.

Race, P. (2006). *The Lecturer's Toolkit: A Practical Guide to Learning, Teaching and Assessment* (3rd ed.). New York: Routledge.

Readman, K., & Allen, B. (2013). Practical Planning and Assessment. South Melbourne: *Journal of the Scholarship of Teaching and Learning*, Vol. 15, No. 5, October, 2015. *Josotl.Indiana.edu*

Oxford University Press.

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. (1987). Specifying and Promulgating Achievement Standards, 191.

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment and Evaluation in Higher Education*, 30 (2), 175-194. doi: 10.1080/0260293042000264262

Sadler, D. R. (2007). Perils in the Meticulous Specification of Goals and Assessment Criteria. *Assessment in Education: Principles, Policy & Practice*, 14 (3), 387-392.

Sadler, D. R. (2009). Grade integrity and the representation of academic achievement. *Studies in Higher Education*, 34 (7), 807-826. doi: 10.1080/03075070802706553

Sadler, D. R. (2010). Fidelity as a precondition for integrity in grading academic achievement. *Assessment & Evaluation in Higher Education*, 35 (6), 727-743. doi: 10.1080/02602930902977756

Sadler, D. R. (2013). Assuring academic achievement standards: from moderation to calibration. *Assessment in Education: Principles, Policy & Practice*, 20 (1), 5-19. doi: 10.1080/0969594X.2012.714742

Tesch, R. (1990). *Qualitative Research: Analysis Types & Software Tools*. London: Falmer Press.