Perspectives on Rubric Design and Construction

This concluding chapter shifts from implementing rubrics in different tertiary learning contexts to designing and constructing rubrics. It takes a more pragmatic perspective by outlining a process for constructing holistic marking rubrics to align course learning outcomes with the assessment intentions. In the first iteration, we construct a rubric using the more familiar matrix format before translating the information into the continuum model outlined by Grainger in chapter 4. We then compare these two different formats for rubric design in terms of their constraining and enabling characteristics when it comes to employing them for making valid judgments about student learning.

Notice that the emphasis in this chapter is rubrics used for evaluating student work rather than the formative use of rubrics to enhance student learning. There is an underlying assumption that any rubric should enhance student achievements, however this is not the focus of the information provided here. To design rubrics for the purpose of making valid judgments, the process must be part of the conceptualisation of the entire assessment instrument. Designing the instrument and the rubric simultaneously improves alignment between the course learning outcomes, the assessment intentions and the marking criteria (Weir, 2009). So, the focus is on the conceptualisation phase of assessment design and this means that the primary audience of these specific texts are the assessors. On completion of this phase the rubric can be altered or adjusted for purposes other than making judgements that target the student audience.

The process of creating a rubric described here sources ideas and practices proposed by acknowledged scholars and researchers who focus on rubric design and construction such as Brookhart (2013; 2018), Dawson (2017) and Reddy (2011). The process also draws on policy guidelines published by statutory educational authorities and it is underpinned by the authors' years of experience with criterion-referenced, standards-based approaches to assessment in both secondary and tertiary educational contexts. The rubric design process takes into account the nuances of higher education policy constraints but it is unashamedly guided by practices that have emerged from the senior schooling sector in Australia (and the state of Queensland in particular) because they have been refining the art of criteria-based assessment for nearly four decades and these principles still apply to the higher education context.

The rubric created by this process complements assessment instruments which we categorise as performance-based and 'authentic'. According to Sadler (2009a), authentic assessment

… accounts for a significant proportion of assessment activity in higher education. They are intended to provide opportunities for students to demonstrate sophisticated cognitive abilities, integration of knowledge, complex problem solving, critical opinion, lateral thinking, and innovative action. (160)

Another way of describing authentic assessment tasks are those that "… encourage deep learning and higher order thinking; offer students flexibility and choice in their assessment response; engage them in authentic tasks that mirror real-life applications of the discipline" (Carless, 2015, 243). To evaluate students' responses to authentic assessment tasks, the rubric design must enable qualitative, holistic judgements to be made about the student's achievement of the course learning outcomes. This points to an important caveat on any rubric design which is the necessity to design the rubric at the same time as the authentic assessment instrument. This will enhance the construct validity of the instrument by ensuring alignment between the course learning outcomes (or curriculum intent) with the assessment requirements and the criteria for marking. How to achieve educational alignment between course learning outcomes and the assessment instrument is briefly explained in the next section before the focus moves to rubric construction.

Holistic Rubric Design for Educational Alignment

It is widely acknowledged that alignment of the three educational "message systems" (Bernstein, 1990) - curriculum, pedagogy and assessment - is necessary for effective learning to occur. For higher education contexts Biggs (1996) coined the term 'constructive alignment' which he describes as an outcomes-based approach to teaching in which the course learning outcomes are defined before considering the teaching and assessment methods.

Another way of achieving educational alignment is via a process of backward mapping (see Weir, 2009; 2019) where, like Biggs we start with the course learning outcomes but then design the assessment instrument before devising the teaching and learning activities. We advocate that devising the assessment first guarantees 'constructive' alignment because the summative assessment instrument represents the destination of part, or all of the course of study, and only after this is established can the necessary teaching and learning experiences be clearly conceived.

Backward mapping is the adaptation of Wiggins and McTighe's (2005) three-step curriculum planning process termed "Backward Design". An extra step at the end of the process has been added so that formative assessment opportunities to monitor student learning progress are included in the plan. The four-step process of backward mapping is outlined below:

Step 1. Ensure course learning outcomes make explicit the cognitive processes involved
Step 2. Devise assessment instruments that will elicit evidence of those learning outcomes
Step 3. Plan the teaching and learning activities
Step 4. Determine key points for monitoring student learning and providing feedback

Rubrics can be adapted for this fourth step in the planning process and for formative learning purposes to enhance student achievement. However, the most important role for rubrics in the backward mapping process is for the marking criteria to encapsulate the course learning outcomes so they align. The way to do this is described by Biggs (2014) as a process of taking the cognitive verbs used to construct the course learning outcomes and then make sure the assessment task instructions contains the same verb as do the marking criteria. This signifies the importance of the language choices made when constructing a rubric so that there is no discourse slippage between the course learning outcomes and the assessment instrument. In the section that follows we demonstrate this translation from the course learning outcomes to task specific, holistic criteria as we outline our approach to rubric design and construction.

Designing holistic marking rubrics

This rubric construction process presented in this chapter will initially use the more familiar matrix format (illustrated below) to present the criteria and standards or levels of learning achievement.

STANDARDS	Outstanding	Very high	High	Satisfactory	Fail
CRITERION					
CRITERION					
CRITERION					

Figure 10.1 Matrix format

Note that the standards are deliberately listed from highest to lowest so that the highest standard of achievement is viewed first, rather than reading what constitutes a fail. The number and titles of the standards listed may differ according to institutional policies. The descriptors of quality expected for each standard and each criterion are inscribed into the remaining matric cells.

Once constructed we transfer the same information into the continuum format which is also referred to as the Continua Model of a Guide to Making Judgements (or GTMJ). This is to demonstrate how different rubric formats can change the message they represent about assessment and learning, and potentially remove some possible constraints they create when making holistic judgements.

Determining the criteria

Two widely acknowledged characteristics of marking criteria are that they must be explicit and task specific (Jonsson & Svingby, 2007; Bloxham, den-Outer, Hudson & Price, 2015; Worth; 2014). Explicit criteria clearly describe the assessable elements of the task in terms of WHAT skills or knowledge are being evaluated. They index what learning is valued in the assessment instrument. Criteria that are task-specific highlight WHERE the learning achievement is observable or identifiable in the students' assessment response.

The first step in constructing a rubric is to establish what learning is being assessed by the task. In other words, what are the assessable elements of the task that are generating evidence of student learning. According to Readman and Allen (2013) criteria are qualities by which students' assessment responses are being judged. They are the properties, characteristics or dimensions by which student performance is appraised. Their function is to communicate effectively to students what they should be learning and the basis on which teachers will assess their work (Wiliam, 2018). When determining the criteria for a task they must embody the curriculum intent and so what is being assessed reflects the course learning outcomes. Having established what is being assessed, the next step is to ensure the criteria signify where the assessable elements can be observed in students' work.

For the purpose of demonstrating our approach to constructing rubrics, we use an authentic example recently implemented in a foundational teacher education course about the relationship between curriculum, pedagogy and assessment and how they affect learning. We have simplified the task for the purpose of demonstrating constructive alignment.

THE ASSESSMENT INSTRUMENT

Task instructions

Students will analyse the Australian Curriculum by comparing two different learning areas in terms of their rationale, knowledge organisation and prescription of content. They are to record their findings in a template (provided) and then use these to write an explanation of the implications of any significant differences between the learning areas for classroom teaching and learning.

Course learning outcomes being assessed:

- Read and interpret the Australian Curriculum by comparing two different learning areas
- Understand the relationship between curriculum and pedagogy
- Communicate curriculum knowledge and ideas applying educational terminology

CRITERIA	WHAT	WHERE
The comparison between the two learning areas exhibits an ability to read and interpret the Australian Curriculum.	Read and interpret the curriculum	The findings from the comparison which are recorded in the template
The explanation of the implications for teaching and learning show understanding of the relationship between curriculum and pedagogy.	Understands the relationship between curriculum and pedagogy	In the written explanation
The language and structure of the written response is evidence of an ability to communicate curriculum understanding.	Communication	Language and structure of the written response.

Table 10.1 Establishing criteria for the curriculum task

Other considerations for determining criteria

The first consideration is the number of criteria in the rubric. Brookhart's (2018) review of rubric in higher education reveals the most common number of criteria in a rubric is three to five. Our anecdotal evidence and experience also suggest there are no hard and fast rules about the number of criteria, however beyond four or five the rubric becomes too 'fine-grained' rendering it unwieldy and unworkable. That is, its utility in making an holistic judgement is undermined by the complexity of the rubric created by having too many criteria.

A second consideration is that the criteria are sufficiently different and independent from each other. Humphry and Heldsinger (2014) emphasise the need to avoid similarities between criteria to ensure

that there are clear distinctions between the elements of the task that are assessable. They further qualify this statement by claiming,

> We do not mean to imply that criteria need to be wholly or even largely independent of each other. Criteria may be mutually related by virtue of their reference to the common construct. Nevertheless, given the aim of using a rubric, it is desirable for criteria to contain descriptions of performances free of obvious overlap or redundancy to allow raters to focus on distinctive and complementary aspects of students' performances, and to capture individual differences in each aspect within ratings. (256)

A final consideration is the specificity or 'grain size' of each criterion. Constructing criteria is a balancing act between capturing the assessable element of the task without listing specific analytic details that narrow its interpretation. One way to approach this balancing act is to 'chunk' the assessable elements together, so that each has a similar weight or value. This is essential for making holistic judgements and it therefore precludes any arbitrary quantitative weightings from inclusion in the rubric. Arbitrary weighting as percentages introduce bias to the judgement through the calculations required to determine the final grade/mark. Any concerns about the lack of specificity by educators can be alleviated by the requirement that these are further unpacked with the learners to establish a shared understanding of the assessment expectations (Black and Wiliam, 2009).

Determining the standards descriptors

The standards in a rubric are levels of achievement that represent qualitative differences in performance. The number of standards will vary according to institutional assessment policies and whether the assessment is graded or non-graded, for example: PASS/FAIL. Standards descriptors are behavioural statements that succinctly convey the expected quality of student response to each criterion for each standard. They describe the learning behaviour assessors are looking for as evidence of student performance. They index the quality or degree of evidence of learning expected for each criterion. These need to be as explicit as possible to avoid confusion for both learners and assessors.

Standard descriptors are so important in conveying the quality of work expected that their construction requires careful attention. Research indicates that the language used to construct standard descriptors is one of the most challenging aspects of rubric design (Tierney and Simon, 2004; Moni, Beswick and Moni, 2005). This is to avoid misinterpretation of the text by the intended audiences (Reddy & Andrade, 2010). One of the reasons standards descriptors are so difficult to write is because the assessor is trying to predict different levels of quality in students' assessment responses. Until the assessment task is implemented, a more consistent delivery of the GTPA curriculum and increased reliability of the marking. The intention here is limited certainty that that the desired learning behaviour has been accurately captured in the standards descriptors. However, future iterations could generate different degrees of quality depending on the student cohort. Thus, from our perspective, rubric design and construction is a perpetual work-in-progress.

In our experience we have found it is best to start with the description for the highest standard for each criterion. Because it is easier to work backwards, once the highest level of quality is captured. The next descriptor to write is the pass standard, because that establishes the threshold level of quality that constitutes the difference between a pass and a fail. After that, the matrix can be completed by choosing discerning qualifiers that describe different levels of quality expected for the remaining standards. At the end of this chapter we provide a useful resource for writing standards descriptors: a series of qualifiers listed beneath five levels of learning achievement.

Having explained our perspective on designing rubrics for holistic judgments, we now present the completed matrix from the authentic assessment example cited above.

STANDARDS/ CRITERIA	Outstanding	Very high	High	Satisfactory	Fail
The comparison between the two learning areas exhibits an ability to read and interpret the Australian Curriculum.	The comparison accurately identifies and clearly describes the key differences between each learning area.	The comparison identifies and clearly describes key differences between each learning area.	The comparison identifies differences between each learning area which may not be clearly described.	A simple comparison that identifies minimal differences that may not be clearly described.	A cursory comparison that demonstrates limited capacity to read and interpret the curriculum.
The explanation of the implications for teaching and learning show understanding of the relationship between curriculum and pedagogy	Insightful implications for teaching and learning are discussed with clear reasoning for each key difference.	Credible implications for teaching and learning are discussed with adequate reasoning for each key difference.		Implications are presented that may not be credible or are poorly explained with limited or no reasoning for each difference.	The discussion demonstrates a lack of basic understanding of the relationship between curriculum and pedagogy.
The language and structure of the written response is evidence of an ability to communicate curriculum understanding.	Publishable quality and sophisticated use of educational terminology.	The text is coherent with a logical structure. Some minor errors in spelling or grammar and effective use of educational terminology.	The text is mostly well organized, but some parts may be difficult to interpret. Sporadic use of educational terminology and some errors present.	The text shows some evidence of organization, but some aspects may be difficult to interpret. Sporadic or improper application of educational terminology and obvious errors present.	The text demonstrates limited ability to communicate curriculum understanding.

Figure 10.2 Standard (matrix) rubric for curriculum task

Now the same information from the matrix format will be translated into the continuum model (shown in Figure 10.3 below), and then we explain how the change in format represents a different message about assessment and learning, and we examine the implication of removing the grid for making holistic judgments.

Outstanding	Very High	High	Pass	Fail

Curriculum Analysis
The comparison of between the two learning areas exhibits an ability to interpret the Australian Curriculum.

The comparison accurately identifies all the key differences between each learning area.	The comparison identifies and clearly describes key differences between each learning area.		A simple comparison that identifies minimal differences that may not be clearly described.	A limited interpretation of the differences between the learning area.

Curriculum implications
The implications for teaching and learning outlined in the response show understanding of the relationship between curriculum and pedagogy.

Insightful implications for teaching and learning are discussed with clear reasoning for each key difference.	Credible implications for teaching and learning are discussed with adequate reasoning for each key difference.		Implications are presented that may not be credible or are poorly explained with limited or no reasoning for each difference.	The discussion demonstrates a lack of basic understanding of the relationship between curriculum and pedagogy.

Communication of curriculum
The language and structure of the written response is evidence of an ability to communicate curriculum understanding.

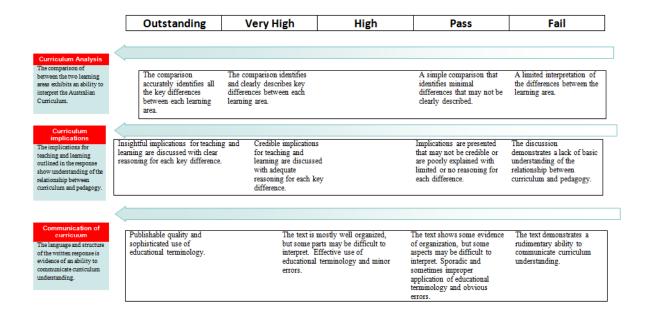|				
Publishable quality and sophisticated use of educational terminology.		The text is mostly well organized, but some parts may be difficult to interpret. Effective use of educational terminology and minor errors.	The text shows some evidence of organization, but some aspects may be difficult to interpret. Sporadic and sometimes improper application of educational terminology and obvious errors.	The text demonstrates a rudimentary ability to communicate curriculum understanding.

Figure 10.3 GTMJ for the curriculum task

There are three distinguishing features of the continuum model which have significant implications for the construction and utility of the rubric. These are the insertion of the arrows, the removal of the grid and the 'nestedness' of the standards descriptors. The arrows represent learning as a continuum, rather than something that happens in discrete quantum leaps. Furthermore, extending the arrows beyond the descriptor signifies that in some instances student achievement may exceed expectation (Grainger and Weir, 2015) and it is an acknowledgement that that the standards descriptors cannot always accurately predict the highest achievable quality.

The removal of the grid reduces the perceived constraint that can sometimes force a judgement into a single cell of the matrix. It also prevents the indecisive marker from ticking the line between two standards, or from suggesting that there is a gradation of quality within the box. The validity of the judgment is questioned if latent quality labels are applied to the judgment. Without the grid, the student learning can be judged anywhere along the continuum as a region of quality rather than a fixed point (Grainger and Weir, 2015). The removal of the grid also means that the standards descriptors for each criterion can be positioned at different points along the continuum to signify relative difficulty in place of artificially weighting each criterion. For example, in figure 10.3 the highest standard for the middle criterion is positioned to the left of the others due to the cognitive complexity demanded by this criterion.

The remaining feature is in the standards descriptors themselves which are constructed by applying the notion of 'nestedness'. This feature means that the defining degree of quality is only specified when a discernible difference is articulated to distinguish a difference between two standards (Grainger and Weir, 2015). This is illustrated in figure 10.3, when the standards descriptors are read

from right to left. So visually the GTMJ differs with a continuum of learning approach with arrows and no boxes that confine the descriptors to a set grid. The features are still representative of a rubric but portray a more fluid perspective on assessment criteria so making judgements is not just an exercise in "*ticking the boxes*".

Regular and collaborative review ongoing process

In concluding this chapter, it is worth revisiting the notion of rubric construction as an iterative and ongoing process. One of the design elements considered above and supported with literature is the iterative and ongoing review of the rubric. The design of a rubric is ongoing as the feedback informs which aspects of the marking criteria were explicit and where an aspect of the marking criteria may not have been as clear, or responses were not as expected. Gibson (2013, as cited in Bharuthram, 2015) believes a rubric is not designed as a one off, but fluid process. Thus, the rubric is adapted based on the learners' responses and the experience of the assessors and any changes should assist in making the rubric a more a valid and reliable marking tool. Bharuthram, (2015, 425) states "that the lecturer should be willing to modify the rubric even during the grading process in order to cater for unanticipated responses from students." We are not advocating the modification during an assessment event as this would undermine the construct validity of the assessment instrument. We are however suggesting that a review of the rubric is normally necessary in light of what students submit in their assessment responses.

Reflective practice is what professional educators do and so the assessment (and rubric) review process can play a significant role in informing teaching practice and contribute to the redevelopment of curriculum. That is, students' assessment achievements can often inform when the curriculum or pedagogy needs to be adjusted to ensure salient points are clearer. In addition, the range of student ability and cultural diversity of the cohort needs to be factored into rubric design, to ensure constructive alignment. Therefore, the rubric needs to be reviewed as appropriate to ensure it remains relevant to the students' experience of learning and assessment.

In summary this chapter has outlined an approach to constructing rubrics for authentic assessment and for making qualitative holistic judgements. We highlighted that this can facilitate constructive alignment between the course curriculum and assessment. There are a number of key messages for academic teaching staff to take away from this chapter. Firstly, how to determine clear criteria that accurately portrays what is valued in the assessment. The second key message is about the construction of the standards descriptors to accurately capture the quality or degree of learning expected and we note that this is quite a challenging aspect of rubric construction and best undertaken collaboratively. Based on this information we demonstrated how a completed rubric can look in practice. Then we compared two different formats and examined their implications for interpreting rubrics and for making holistic judgements. The final key message is that rubric construction is not a one-off process and that these texts are always a work in progress. The practical perspective taken here should provide enough guidance to remove some of the perceived difficulty of rubric construction.

Appendix

Outstanding	High	Satisfactory	Marginal	Fail
Comprehensive Insightful Proficient Discerning Well-reasoned Clear Perceptive Controlled Skilful Accurate Significant Well-justified	Thorough Thoughtful Logical Coherent Effective Logical Purposeful Informed Accurate Proficient	Suitable Competent Relevant Credible Sound Appropriate Functional	Narrow Variable Disjointed Superficial	Rudimentary Minimal Unclear Cursory Vague

Table 10.2 Qualifiers for writing standards descriptors
(Adapted from: https://www.qcaa.qld.edu.au/downloads/p_10/qcar_standards.ppt)