

Guidelines for Building Marking Rubrics

Why should we take the time to build rubrics?

A rubric is a framework that lists specific criteria for grading academic work. Additionally, the rubric will contain descriptions of different levels of performance. With the introduction of the new degree classification based on a 70:30 (final year %: second year %) algorithm, it is essential that students receive clear guidance about the standards expected for different levels of performance. Well-designed marking rubrics with appropriate criteria and well differentiated descriptions of performance will benefit staff and students alike as they work as reliable measurement tools for assessment, and as effective learning and teaching aids.

Some advantages:

- Rubrics can be used as a self-assessment tool for students to help them plan and check their draft work
- Rubrics will help tutors to identify and focus on the key knowledge, skills and/or attributes to be developed within a module
- Inter-assessor and intra-assessor reliability will be improved
- The process of marking can be faster and it will be easier to reliably allocate marks and differentiate between students.
- Predetermined criteria with descriptors will be applied to each assessment, making the process more likely to be free from bias
- Students will see the marking process as transparent, reliable and fair and will be less likely to query an assessment result

Step 1: Building the rubric

What are the elements of a marking rubric?

A marking rubric contains the assessment marking criteria, the level scale and the associated level descriptors. Marking criteria will elaborate on the module learning outcome(s) and will provide the finer detail of what is expected in submitted work. The associated descriptors will differentiate between levels of performance.

Level Scale and descriptors

	0-39%	40-49%	50-59%	60-69%	70-79%	80-100%
	fail	3rd	2.2	2.1	1st	High 1st
criterion	descriptor			descriptor		
criterion		descriptor			descriptor	
criterion			descriptor			descriptor

If an assessment strategy includes multiple assessment items of varying methods that measure divergent knowledge, attributes and skills then separate rubrics would need to be built for each item.

Level Scale

Assessment criteria are typically assessed across a continuum of grading levels ranging from fail to 1st class. Due to the large span of grades at either end of the scale (i.e. 0-39 and 70-100), it is good practice to subdivide the levels further to include e.g. a low fail (0-29) and a high 1st (80-100). This provides further guidance for students on the expected level of performance across the continuum. The extended levels can also encourage markers to use the full marking range.

Stepped Marking

Differentiating student work can be challenging. Some markers might quibble over 1 or 2% grade differences on a piece of work e.g. whether to give 66% or 67%, even though this difference could have little impact on the calculated degree classification. Additionally, some markers might prefer to mark to grade boundaries, i.e. awarding just 70% for 1st class work or 60% for 2.1. etc. This practice would make it difficult for students to gain a 1st degree overall within the classification system.

Stepped marking defines a selection of agreed grades for each level to delineate grade boundaries and to help differentiate student work. When marking work at the level of a 2.1 for example, a marker can decide whether the work is worth a high, mid or low 2.1 depending on the degree to which the work meets the associated descriptor.

The example below demonstrates the ‘steps’ between each level. When awarding a 1st level for example, a marker would need to allocate at least 75% to indicate that they were confident that the work was appropriate for this level. This approach would also ensure that more students would gain 1st class degrees when aggregate marks are calculated. *Again, it is good practice to extend the grade levels at either end of the continuum for further clarity and to ensure the full range is used.*

Example of stepped grade levels

Fail	Pass	3 rd	2.2	2.1	1 st
0, 5, 15, 25, 35, 39 %	42%	45, 49%	52, 55, 59%	62, 65, 69%	75, 80, 85, 90, 95, 100%

Note that if a grade of e.g. 69% is awarded against a given criterion, student feedback should indicate that work has **fully** met this grade level but additionally should outline reasons why the higher level has not been achieved. The associated grade level descriptors will help to provide this level of clarity and differentiation.

How do I choose the marking criteria?

When deciding on marking criteria, we need to focus on what we want our students to learn within a given module and how this will be assessed. Revisiting the module learning outcomes provides some initial focus.

A learning outcome typically contains an **action verb** and a **context** for this verb. Both of these elements will be considered within the associated marking criteria

*E.g. Students will be able to **discuss the impact of socioeconomic status in relation to access to higher education.***

In this example, an assessment must enable a student to demonstrate their ability to build a discussion and their ability to demonstrate knowledge of HE (through the discussion).

For action verbs e.g. discuss, evaluate, reflect, plan, present, design etc., we need to consider what is involved in these processes. What are the key elements of a discussion for example, within a given context? Criteria will need to cover these elements, including the broader subject knowledge.

Action verbs e.g. the ability to 'discuss' can be conveyed in a variety of different ways such as written, verbal, multimedia etc. The assessment method, which is aligned closely to the learning outcome(s), will influence the selection of marking criteria. If an assessment is a videocast for example, then we need to consider building in criteria that relate to video production, communication and presenting skills, as well as knowledge of the subject area and the ability to build a discussion.

Selecting a videocast as an assessment method prompts us to think about the inherent skills involved in this method and in turn, could prompt us to revisit our learning outcomes to include an outcome relating to communication, presentation or digital skills:

E.g. students will be able to communicate effectively with different audiences using a range of media

Additionally, generic attributes are often included within criteria; structure and referencing are typical examples. These generic criteria would be given less weighting than criteria associated more closely with learning outcomes (see section on weighting).

So essentially, the rubric needs to be specific to the subject context *and* the assessment method (aligned to the active verb) in order to measure the learning outcomes effectively. The associated marking rubric will need to cover the range of knowledge and skills that is being developed through the assessment within the list of criteria. These criteria must be able to support descriptions along a continuum of quality. Criteria and descriptors must be measurable, distinct and clearly described.

Weighting the criteria and calculating grades

The rubric will usually contain a mix of outcome/assessment – specific criteria as well as criteria relating to generic attributes e.g. structure, referencing etc. Weighting the criteria will help students to see the value you are placing on each element of their work.

There is no specific guidance to help quantify this weighting and so academic judgment is required. Nevertheless, criteria that are associated closely with the module learning outcome(s) will be more heavily weighted than criteria associated with generic attributes (structure/referencing etc.). Criteria that measure greater degrees of cognitive effort will be weighted more heavily than criteria that are less complex.

The varying weightings must sum to 100% of the whole.

A simple calculation is then used to generate an overall percentage grade for the piece of work. In the example below, the marker could decide that criterion 1 has *just* been met and could award 42% for this criterion. The criterion applies the weighting (40% weighting in this example) so $0.4 \times 42 = 16.8$ marks for this criterion. The total marks are then added and rounded up to give the overall grade.

Example grading calculation

Criteria (100%)	0-39	40-49	50-59	60-69	70-79	80-100
Criterion 1 40%		0.4 x 42 = 16.8 marks				
Criterion 2 30%			0.3 x 55 = 16.5 marks			
Criterion 3 20%			0.2 x 59 = 11.8 marks			
Criterion 4 10%				0.1 x 65 = 6.5marks		

Total marks = 51.6 – 52% overall grade

Sharing the marking rubric within student feedback will help students to see where marks have been awarded and will make the whole process much more transparent.

How do I gauge the level of the rubric?

Assessments are set at different levels of learning and so marking rubrics must also reflect the appropriate higher education level of the course and module. [The Framework for HE Qualifications](#) (QAA, 2014) contains qualification descriptor sets that describe the threshold academic standard for qualifications in terms of levels of knowledge, understanding and abilities. This will be used as a reference point when creating programme and module learning outcomes but it can also be used to help determine threshold levels when building marking rubrics. Alternative frameworks are also useful such as Subject Benchmarks, the EwNI Generic Credit Level descriptors or the [Credit Level Descriptors for Higher Education](#) (SEEC, 2016).

An example of a descriptor from the knowledge and understanding category, at level 4 of the SEEC framework is outlined in the table below:

Knowledge and Understanding	
Knowledge and understanding	Has a broad understanding of the knowledge base and its terminology or discourse. Appreciates that areas of this knowledge base are open to ongoing debate and reformulation.

Adapted from SEEC, 2016

When building a marking rubric, consider the threshold level of performance i.e. the pass level, first. We should be satisfied that work at this grade (just) meets the required level. We can draw from the relevant frameworks to help build the descriptor at this threshold pass level.

For example, if we were building a threshold pass descriptor at level 4 relating to subject/context knowledge and understanding, we might indicate that work demonstrates a *'basic understanding of the knowledge base although terminology sometimes lacks precision and some concepts are a little biased or outdated...'*. This clearly maps to the SEEC descriptor above but also indicates areas that require further development.

It is typical to include adjectives as discriminators e.g. satisfactory, good, very good etc. and a selection is provided in the table below. However, *do not* simply rely on these alone as they provide no clear guidance about what e.g. 'good evidence of discussion' looks like. Vague descriptors are likely to result in subjective marking.

Each descriptor should provide a clear outline of performance at each level. Once the threshold level descriptor is defined, we can build on this incrementally adding additional attributes but also indicating areas needing further development.

Examples of discriminator adjectives

0-39	40-49	50-59	60-69	70-79	80-100
Lack of... No evidence Very limited... Muddled... Insufficient...	Some evidence... Limited... Satisfactory... Adequate... Moderate...	Clear... Good... Solid... Sound...	Very good... Informed... Thorough... Strong...	Excellent... Effective... Exemplary... Extensive...	Outstanding Innovative Original Creative
Building descriptors incrementally:					
Work does not reach the threshold level	Threshold level achieved. Indicate what is missing or what needs developing	Build on threshold achievement adding additional attributes. Indicate what is missing or what needs developing,	Build on previous description adding additional attributes. Indicate what is missing or what needs developing	Build on previous description	Work at this level surpasses expectations. Acknowledge creativity or unintended outcomes. Work could not be improved at undergraduate level in the time available

The rubric encourages markers to use the full range of grades. Descriptors at the highest grade level should define work that exceeds expectations or demonstrates original thinking perhaps, and can encourage students to excel in their work.

Step 2. Reviewing the draft rubric

The marking rubric is a key assessment tool and should help to ensure stable and consistent results when applied to measure student work. It is important for the whole marking team to review and test a draft rubric to help ensure inter-assessor reliability. Different markers could interpret criteria differently if the criteria and descriptors are vague and poorly defined. Discussions are valuable before, during and after the implementation of a rubric to ensure

that is robust and reliable across the marking team. It is useful for teams review rubrics across a programme(s) to check the quality, consistency and level.

The wording within rubrics must be clear and unambiguous but must not be overly prescriptive. A rubric must be well aligned to the module learning outcomes and to the content delivered/activities within the module.

It is not unusual to revise rubrics after use, as student work will have revealed common pitfalls, examples of good practice or different interpretations that may be incorporated into future versions. The design of a rubric requires continuous adaptation and development.

Step 3: Sharing the Rubric with Students

Marking rubrics are valuable learning tools for students and can be used as part of a formative self- assessment of their work in progress. Rubrics can build assessment literacy as they can help students understand the targets for their learning and the standards of quality for a particular assessment (Reddy & Andrade, 2010). Sharing the marking rubric with students during formative activities will ensure that both the lecturer and the students have the same expectations of the assessment (Cox, Morrison & Braithwaite, 2015) and this can help students achieve higher marks (Rust, Price & O'Donovan, 2003). This transparency should also reduce potential claims of biased or unfair marking.

A Note on Checklists and Rating Scales

The term rubric can be used incorrectly sometimes, and can be confused with checklists or rating scales.

- A rubric will include criteria for students' work *and* descriptions of performance quality
- A checklist is a list of characteristics with a place for marking whether the characteristic is present or absent (yes/no). The checklist could be used for some competency pass/fail assessments to check the existence of an attribute but not its quality. Checklists can also be used as formative tools to enable students to check whether they have included all the elements of an assignment prior to submission.
- A rating scale is a list of characteristics with a place for marking the degree to which each characteristic is displayed. They indicate performance levels (e.g. satisfactory, good, excellent etc.) or frequency (seldom, often, very often etc.) but do not include the descriptions of performance quality. The example below demonstrates a numerical scale that indicates high expectations of performance at one end to low expectations at the other.

Example of a criterion and associated rating scale:

Introduction

Well-structured introduction	5	4	3	2	1	No introduction
------------------------------	---	---	---	---	---	-----------------

Rating scales can be quick to build and seemingly efficient during real-time assessments such as presentations or performances, but they are open to subjective marking. Using the example above, there is no indication of what a ‘well-structured introduction’ might look like at a given level of learning. Additionally, there is no indication of a threshold level of performance. One marker’s subjective view of satisfactory might be a 3, but for another marker it might be a 2. And what does ‘satisfactory’ even look like?

One could argue that rating scales could be used when double marking is applied to an assessment (e.g. during a real-time performance), *and* if both assessors discuss and agree marks after the assessment. However, rating scales have limited value as formative tools for students as they provide no information that will move learning forward. What does the student need to do to move from a 3 or a satisfactory up to a 5 or excellent?

If you are currently using checklists or ratings scales, there is scope to review these to determine whether they can be adapted into rubrics. There is even potential to involve students at this stage as they could contribute in defining the level descriptors for each criterion. This could be another method of building their assessment literacy.

References:

Brookhart, S.M. (2013) *How to create and use rubrics for formative assessment and grading*. Alexandria, Virginia, USA. Association for Supervision & Curriculum Development

Cox, G. C., Brathwaite, B. H., & Morrison, J. (2015). *The Rubric: An assessment tool to guide students and markers*. *Advances in Higher Education*, 149-163.

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

Rust, C., Price, M., & O'Donovan, B. (2003). *Improving students' learning by developing their understanding of assessment criteria and processes*. *Assessment & Evaluation in Higher Education*, 28(2), 147-164.