3.4

Assessment Rubrics: Design, Development and Implementation

Folashade Akinmolayan Taiwo[a], and Erik Vellinga[b]

[a] Queen Mary University of London, United Kingdom, ORCID: 0009-0008-6830-1455;
[b] Delft University of Technology, the Netherlands, ORCID: 0009-0006-8418-6317.

Keywords: *rubric, feedback, assessment tools, alignment, module design, learning outcomes*

This chapter examines the design, development, and implementation of assessment rubrics as practical tools for evaluating transferable skills in engineering education. Transferable skills are complex, context-dependent, and often assessed through authentic tasks such as projects, laboratory work, teamwork, and professional communication, making consistent and transparent assessment challenging. Rubrics address this challenge by articulating expectations through explicit criteria and performance standards, thereby supporting fairer and more consistent judgement. The chapter introduces foundational concepts underpinning effective rubric use, including evaluative criteria, performance levels, and scoring strategies, and distinguishes between commonly used rubric types such as holistic and analytic rubrics. It discusses how rubrics can be applied for formative and summative assessment purposes, highlighting their roles in supporting feedback, learning, and assessment reliability. This chapter also offers a step-by-step, practitioner-oriented guide to rubric development. This guide addresses purpose clarification, alignment with learning outcomes, criteria formulation, performance-level description, and iterative refinement through feedback and assessor calibration. Examples drawn from engineering education illustrate how rubrics can be designed and implemented to assess transferable skills within authentic assessment contexts.

DOI: 10.59490/mt.266.79.

1 Introduction

Assessing a transferable competency or skill is difficult. These skills are often complex and context-dependent and they may require forms of assessment other than exams or similar tests. Because of their complexity, expectations of how a task involving these skills is executed well may differ between a student and an assessor - or even within a group of assessors themselves. It requires some effort to develop a framework that supports more reliable and less subjective assessment of these skills. Rubrics address this need by translating complex performance expectations into explicit, structured criteria.

A rubric can be defined as a tool that articulates expectations for student work by listing assessment criteria and describing levels of performance along a continuum of quality (Brookhart, 2018). In the context of engineering education, rubrics are commonly used to assess complex and authentic tasks such as design projects, laboratory reports, teamwork, and professional communication.

Rubrics are also particularly well-suited for assessing transferable skills. These skills are not always directly observable and are demonstrated differently across contexts. Therefore, assessors need a common reference point to judge performance consistently and transparently. By making expectations explicit, rubrics enable students to better understand and improve their skills. The clarity of a rubric promotes learning: in an analysis of studies addressing rubrics and performance, Brookhart and Chen (2015) concluded that rubric use is generally associated with increased student achievement.

Second, as a feedback tool, a rubric can be used by experts to assess the quality of a student's work, but also by students themselves to assess their own work. Regarding the latter, well-designed rubrics improve self-regulated learning, as students use rubrics as reference point to assess their work in progress (Panadero & Jonsson, 2013).

Third, as a grading tool, rubrics are commonly associated with improved inter-rater reliability (consensus between assessor) and intra-rater reliability (consistency from an individual assessor). Jonsson and Svingby (2007) found that the use of rubrics can support more consistent scoring, particularly when agreement is measured across adjacent performance levels rather than exact scores.

Despite their many benefits, the use of rubrics has been criticised (see Panadero & Jonsson, 2020, for an overview of criticisms). For one, rubrics are alleged to fragment complex skills into smaller chunks, essentially asking assessors to focus on "*the parts, rather than the whole*" (Panadero & Jonsson, 2020, p. 8). Similarly, sharing explicit criteria with students may steer students to 'criteria compliance'; and their lack of expertise may make it difficult for students to understand the criteria (Panadero & Jonsson, 2020).

Many criticisms and problems associated with rubrics boil down to poor rubric construction (Oakleaf, 2009). When the wording in a rubric is too general, for example,

it may not be clear to rubric users how to interpret specific criteria. In turn, this may hinder student learning or reliable assessments. On the other hand, when the wording in a rubric is too specific, a rubric may indeed lead to criteria compliance. This means that the rubric is tailored to a very specific assignment, which will not help students in learning how to apply their skills in different settings. In fact, a very explicit task-specific rubric may restrict students to conformity, excluding an innovative 'out of the box' response to the task (McKenna, 2007).

Developing a workable rubric takes time and thought. Rubric development should be informed by the learning outcomes that an assessment task is intended to evaluate. At programme level, learning outcomes articulate the overarching knowledge, skills, and graduate attributes students are expected to develop, while module-level outcomes specify how these are addressed within individual modules. For rubric design, it is sufficient that criteria are explicitly aligned with the relevant module learning outcomes associated with the assessment task.

To make this process easier, this chapter introduces a step-by-step guide to construct a rubric. To do so, Section 2 defines the term and discusses rubric types (holistic or analytic) and their focus (general or task-specific). Then, Section 3 details the role of rubrics in engineering education, detailing their relation to quality standards, assessment procedures, and feedback. In this regard, common pitfalls are discussed that may hinder an optimal use of a rubric. Section 4 shows best practices in developing rubrics, which includes the step-by-step guide in Section 4.1 and a checklist in Section 4.2. Section 5 shows examples of rubrics in engineering education.

2 The role of rubrics

A rubric is an assessment tool used to assess the quality of student work by defining criteria and describing levels of performance in relation to those criteria. In engineering education, rubrics are particularly valuable for assessing transferable skills within complex, authentic tasks that involve judgement and multiple assessors. In practice, rubrics may serve multiple purposes: they may be task-specific or used across a module or programme; applied formatively or summatively; and used by tutors, peers, or students themselves for self-assessment. Clarifying the intended purpose of a rubric is a critical first step in its design.

By making assessment expectations clear, rubrics help students understand performance standards, provide structured and actionable feedback, improve the efficiency and consistency of marking, and may help reduce differences in outcomes between student groups by making assessment decisions more transparent and less subjective.

2.1 Structure of effective rubrics

Rubrics are visualised as tables composed of rows, columns, and cells and usually include three essential features: (1) criteria for evaluation, which specify the aspects of performance being judged; (2) performance levels, which describe gradations of quality for each criterion; and (3) a scoring strategy, which explains how judgements

are translated into feedback or grades. Criteria define 'what' is being assessed, while performance levels describe 'how well' it has been demonstrated.

In Fig. 3.4.1, the rows of the table represent assessment criteria aligned with the intended learning outcomes, while the columns represent different levels of performance (for example, from inadequate to excellent). The cells at the intersection of rows and columns contain qualitative descriptions that define what performance looks like at each level for each criterion, and, where appropriate, indicate associated scores or weightings. This structured visualisation enables both assessors and students to interpret expectations consistently and transparently.

For a rubric to be effective, it is important that the basics are in place, understood, and aligned (Dawson, 2017). The development of assessment criteria should be grounded in well-aligned programme-level and module learning outcomes, allowing educators to select assessment tasks that meaningfully support student development and demonstrate achievement of those outcomes. In engineering education, such tasks should be designed to assess higher-order thinking skills, such as analysis, synthesis, and evaluation, rather than focusing solely on the recall of information (Biggs & Tang, 2011). In doing so, assessments provide opportunities for students to demonstrate progress toward institutional graduate attributes.

Assessment criteria should clearly communicate the specific aspects of the deliverable that will be evaluated in relation to the intended learning outcomes. Developing these criteria and the associated qualitative performance-level descriptors benefits from collaboration among educators to ensure shared standards and consistent interpretation. In addition, involving students in the design of rubrics can further enhance transparency, foster a shared understanding of expectations, and support deeper engagement with the assessment process (Andrade & Du, 2005).

How performance levels are defined may depend on the use of a rubric. Rubrics designed for grading often assign a weight to particular criterion or performance level as part of a scoring strategy (Popham, 1997). When a rubric is used summatively, performance levels may be defined with a numerical descriptor (e.g., '*1 point*' and '*4 points*', as in Fig. 3.4.2). Rubrics can also be used in formative assessments, in which a verbal descriptor may make more sense. These descriptors can be generally evaluative (e.g., *'poor', 'sufficient'*, and '*good*'); they can qualify the student (e.g., *'novice', 'apprentice'*, and *'expert'*), qualify their process (e.g., '*getting started', 'well on your way',* and '*as good as done*'); there are indeed many other possibilities.

2.2 Types of rubrics in engineering education

In engineering education, two main rubric types are commonly used: holistic rubrics and analytic rubrics. The choice between these depends on the nature of the learning outcomes, the assessment task, and the type of feedback required.

Holistic rubrics

Holistic rubrics are often used when one complex skill or multiple related skills are assessed as one. For example, if students of maritime technology develop a ship in a

certain project, the main question in the assessment may be: does it actually do what it is required to do? Despite the complex nature of such a project — which includes skills such as mathematical modelling, spatial reasoning and applying structural mechanics — the entire project is assessed in one single criterion.

As such, a holistic rubric in this project may include the ship's functionality, its design, and other features, but these features are assessed as one. A prototypical holistic rubric can be found in Fig. 3.4.1 and its pros and cons in Table 3.4.1. Holistic rubrics are well-suited for summative assessment and tasks focusing on overall performance, such as examinations.

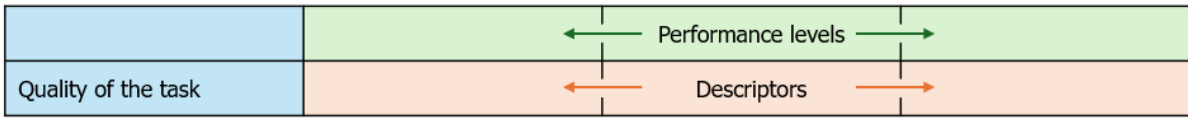

Figure 3.4.1: Example of a holistic rubric illustrating the relationship between criterion, standards, and descriptors (Own Work)

Table 3.4.1: Pros and cons of holistic rubrics

Pros	*Cons*
Fast and easy to apply	Offers minimal specific feedback on components
It is quicker to mark student work (fewer decisions for markers)	Unable to assign importance to individual criteria
Can be consistently applied across graders once trained	It can be difficult to understand how each criterion is weighted
The final mark represents a holistic judgement of the student's progress against all specified learning outcomes.	

Analytic rubrics

An analytic rubric includes multiple specific criteria which together form the assessment of a task. For example, when the same students of maritime technology must present their ship in an oral presentation or a written report, these tasks may be assessed on clarity, on style, storyline, and so forth. Each of these individual criteria is assessed separately as part of the task.

A prototypical analytic rubric format can be seen in Fig. 3.4.2 with criteria being the rows in the table, performance levels the columns, and performance level descriptions the individual cells. The performance level descriptions clarify what counts as a specific level in a specific criterion.

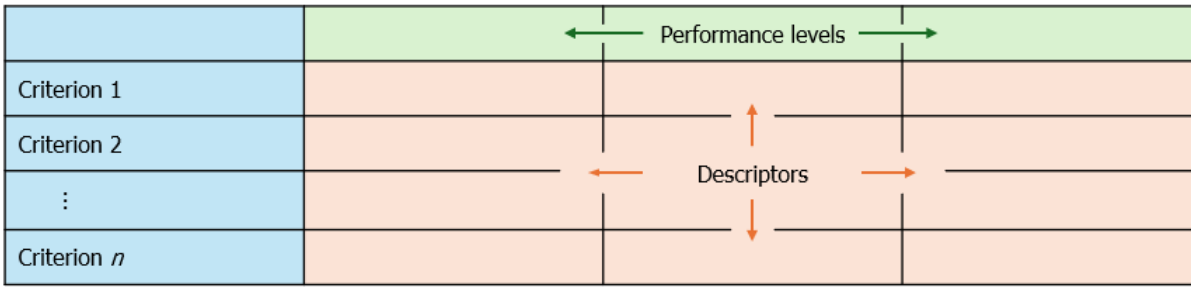

Figure 3.4.2: Example of an analytic rubric illustrating the relationship between criterion, standards, and descriptors (Own Work)

Analytic rubric with criteria that vary in numerical weight

A weighted rubric assigns a specific percentage or a higher point value to certain criteria to reflect their importance, e.g. with criterion A contributing 10% to the grade, criterion B 25%, criterion C 20%, etc. Table 3.4.2. shows the pros and cons of analytic rubrics with criteria that vary in weight. These types of rubrics are well-suited for summative assessment and tasks focusing on overall performance, such as reports, presentations etc.

Table 3.4.2. Pros and cons of analytic rubrics that vary in weight

Pros	*Cons*
It is transparent how the final grade is reached	Time-consuming to design and use
Allows for easier assessment of measurable features	It can place disproportionate emphasis on things that are easy to measure, like grammar, spelling, formatting, or citation style. ('halo effect', see Sadler, 2009)
Able to weight criteria to emphasise key learning outcomes	It can disadvantage multilingual students or those with different academic writing backgrounds, even if they meet core learning outcomes
It is easier to assign the same grade to work of equivalent quality	

Analytic rubric with criteria that have the same numerical weight

With this rubric, every criterion carries the same mathematical weight, e.g., five criteria of 20% each. The pros and cons of such a rubric are shown in Table 4.4.3. These rubrics are well-suited for summative assessment and tasks focusing on overall performance, such as peer assessment, self-reflection etc.

Table 3.4.3: Pros and cons of analytic rubrics with equal weight

Pros	*Cons*
The final grade reflects the work as a whole	Time-consuming to design and use
Provides specific feedback across criteria	Can be problematic if some aspects of the work are more important than others, because all criteria contribute equally to the final mark
Focus is placed on student performance, not scores for each criterion	Too many unweighted criteria may dilute focus or over-partition of tasks

2.3 General and task-specific rubrics

Besides different types, rubrics can also differ from each other when it comes to focus: some rubrics are designed to assess a specific assignment, whereas other rubrics are aimed to skills. This is where the difference between general and task-specific rubrics comes in.

General rubrics assess the performance of a particular skill, such as writing, designing, or solving mathematical problems. By being more general, the same rubric can be used multiple times in a curriculum if the same skill is assessed in different tasks or modules. For example, if a student has to write technical reports in several modules, they might be assessed with the same rubric. In this way, students will be able to understand what it means to be good at a particular skill, and they learn what to expect when performing a task that requires this skill.

A task-specific rubric is geared towards a particular assignment. If we use the same example of students writing a report, a task-specific rubric may include a criterion like 'justification of bio-inspired design'. This criterion is very specific to the assignment, and the rubric therefore cannot easily be used in other contexts. Despite this drawback, task-specific rubrics may make it easier to assess a deliverable, as it allows the assessor to focus on specific elements in a specific task.

In both types, finding the right balance between specific and general is key. If a rubric becomes too specific, they might lead to conformity (McKenna, 2007). For example, if the rubric of a design project details per step of the process what the criteria for quality are, this may hinder students' creativity, and they may falsely assume that all design processes follow the same steps. Additionally, an assessor may not know what to do in an overly specific rubric if they encounter something that was not foreseen by the rubric.

Conversely, if a rubric is too generic, it may lose its value as a didactical instrument. Students may find the rubric too vague and do not know what to do with it. This can

be a problem for assessors too, as a rubric that is too general provides insufficient instruction on how to assess certain behaviour.

3 Context

The use of rubrics in engineering education is supported by a growing body of evidence highlighting their benefits for consistency, transparency, and learning (Cockett & Jackson, 2018; Panadero & Jonsson, 2020). Well-designed rubrics can support fairer marking across assessors, clarify expectations for students, structure formative feedback, increase marking efficiency, and contribute to reducing awarding gaps by limiting subjective judgement. However, these benefits depend on careful alignment between rubric design, assessment purpose, and implementation practices.

For this reason, before developing a rubric in engineering education, educators should address two important questions: (1) what impact is the rubric intended to have; and (2) what is the specific purpose of this rubric? Considering impact involves reflecting on issues such as fairness, reliability among assessors, and the effects of the rubric on student learning and engagement. Purpose, in contrast, relates to how and when the rubric will be used, for example whether it is intended for summative grading, formative feedback, or both. These different purposes may require different design choices; a summative rubric, for instance, may prioritise workability and consistency, whereas a formative rubric may place greater emphasis on explicit guidance and developmental feedback.

3.1 Impact of rubrics

When developing rubrics in engineering education, there are two main stakeholders to take into account: students and (fellow) assessors. A rubric should be clear in order for an assessor to assess the students' work and has to be clear for the students in order to know what is expected from them in terms of quality of their deliverable. To that end, there are certain things a rubric can do, but certainly also things a rubric cannot do.

Most importantly, when rubrics result in a specific score or grade, it can never offer a perfect representation of a skill. Still, when the rubric is well-designed, the score will be a good approximation of that skill. This difference may seem trivial, but many mistakes in developing rubrics can be prevented when priorities are set straight. For example, a common pitfall in making rubrics is to overload them with too many criteria, which may stem from the urge to make a rubric complete instead of effective. A rubric should not be an overview of everything that can potentially go right or wrong in each task.

Equally, when an assessor awards a maximum score to a student based on their scoring rubric (100%, 10 out of 10, etc.), it does not mean that a student has nothing to learn and that their work was perfect. Instead, it means that there was not that much more that reasonably could have been expected of the student given the specific task in a specific context. A high grade acknowledges a student's performance, but they might have scored differently on a different or more complex task.

Ultimately, a rubric serves as an instrument to manage expectations, as if it were an agreement between assessors and assessees on what counts as a good performance of the evaluated skill. In this light, one of the main aims of a rubric is to make it easy to justify a particular assessment — whether it is formatively or summatively — based on observable, relevant and explicit criteria.

A rubric used for grading should also reflect agreement between assessors. Assessing works by a larger student population often requires a team of assessors rather than just one. It would be unfair to students if the same work would receive significantly different grades if they were assessed by different assessors. This means that ideally, when asked to do so, each assessor needs to arrive at the same grade as their peers when they grade the same assignment, to ensure that each grade is fair and not dependent on whoever graded the task. In other words: a rubric should enhance inter-rater reliability.

Research in the inter-rater reliability of rubrics is largely context-specific, focusing on particular rubrics used by particular groups of assessors rather than comparing rubrics with alternative scoring methods. For example, Hansson et al. (2014) reported acceptable levels of inter-rater reliability and agreement when using a rubric to assess scientific writing. However, their findings relate only to that specific rubric and context and cannot be generalised to rubrics as a whole. Similarly, Oakleaf (2009) found that raters could achieve reasonable agreement when using a rubric to assess information literacy, but the study focused on differences related to assessor expertise rather than comparing rubrics with other assessment approaches. Together, these studies suggest that rubrics can support reliable assessment when carefully designed and implemented, but they do not provide evidence that rubrics are more reliable than other scoring methods such as checklists or rating scales.

Jonsson and Svingby (2007) reviewed studies that used rubrics and found that assessors often did not achieve the commonly used benchmark of 70% exact agreement, where two assessors assign precisely the same score to a piece of work. This suggests that perfect consistency between assessors is difficult to achieve, even when rubrics are used. However, they also found that studies measuring adjacent agreement, where assessors' scores differ by no more than one point on the rubric scale, frequently reported agreement rates above 90%. These findings indicate that, although exact agreement is uncommon, rubrics can still support a high level of consistency in assessment decisions.

There are strategies to mitigate problems relating to inter-rater reliability. The main strategy to ensure inter-rater reliability is to discuss assessment with fellow assessors. This can take many forms: a team of assessors can work together to create a rubric, for instance. This allows assessors to discuss what they find important in a skill and what should be prioritised. When there is already a rubric in place, assessors can calibrate scoring (Sadler, 2011). When assessors all assess the same work, compare their assessments, and discuss differences, they do not only reduce differences in their assessment, but they also become more confident in assessing students' work (O'Connell et al., 2015). For new assessors, a rater training may have this effect too.

Other strategies relate to the rubric design. For example, Jonsson and Svingby (2007) note that rubrics with fewer levels generally see higher agreement rates. Additionally, it may benefit assessors to provide explicit scoring guidelines. Examples or benchmarks may help new raters to see how they should use a rubric and how others would have assessed specific works. A potential pitfall in using these examples may be that they become outdated after a while. Generally, it is advisable to update rubrics, scoring guidelines, and examples regularly, to make sure that grading remains valid and reliable.

3.2 Purpose of rubrics

The design and application of a rubric vary on its purpose: rubrics can be used in formative assessment and in summative assessment. Summative rubrics are used to assess a deliverable (e.g., presentation, project, design, text). Typically, a summative rubric results in a grade of some sort, to indicate a student's performance. Formative rubrics are used during a module or in general: the learning process) to identify what goes well already and what needs to be improved.

As formative rubrics are aimed to give feedback on the learning process during a specific task, they need to be clear for students to be able to understand what they can improve and how. As always, the rubric should match learning outcomes and indicate whether students are on track to reach these learning outcomes. For this purpose, formative rubrics are usually very explicit.

Formative rubrics can be employed in a variety of learning activities. For example, students can give each other peer feedback with a formative rubric made by the lecturer. A fun, rather analytic activity is to co-create a rubric with students in class. This allows students to analyse the particular skill they will be learning and set goals for themselves.

A summative rubric should facilitate a valid judgement on a specific task. This judgement should reflect the quality of the performance and, by extension, the student's learning process. In summative rubrics, workability is a key criterion: the rubric should be easy to use for assessors, while still maintaining inter-rater and intra-rater reliability.

Summative rubrics should be shared with students from the start of a module, in order to manage expectations. This way, students can use the rubric throughout their learning process and have a clear understanding of what quality in a particular skill means. After assessing, filled out rubrics should also be shared with students to ensure transparency in grading.

4 Best practices for developing rubrics

Designing effective rubrics is a critical component of fair, transparent, and consistent assessment. Well-constructed rubrics not only support reliable marking but also enhance student learning by clarifying expectations and providing meaningful feedback. This section outlines best practices to guide educators through the process of developing high-quality rubrics; from defining their purpose to refining descriptors

through feedback. The practical steps presented here are grounded in pedagogical principles and aligned with institutional assessment standards.

4.1 Practical steps in rubric development

Step 1. Identify learning outcomes to be assessed

Identify which learning outcomes from the module will be assessed by the specific task or assignment. These outcomes should be clearly listed in the module specification.

Step 2. Define the purpose

First, decide what kind of rubric will be used, such as analytic or holistic, and whether it will give feedback in words (qualitative) or numbers (quantitative). If using numbers, the educator should think about the grade ranges used (for example, below 30%, 30-39%, 40-49%) and the words that describe each level (like 'fail', 'satisfactory', 'good'). It is a good idea to follow the same grade bands and language used in your institution's general assessment guidelines or the commonly used grading standards in your institution if formal guidelines are lacking.

Educators should also decide whether the rubric will be general or task-specific. This distinction influences how criteria and descriptors are written and how the rubric can be reused across assessments. In practice, educators may combine both approaches: using general criteria (such as analysis, communication, or application of engineering principles) while adapting descriptors to reflect the specific task context.

Step 3. Define the assessment criterion

The assessment criteria are the specific dimension of student performance that will be judged. Criteria are the essential features or attributes of student work that relate directly to the learning outcomes. They guide assessors in evaluating performance consistently and transparently and help students understand what is expected of them. Therefore, we recommend educators to:

- *Align with learning outcomes.* Criteria must directly reflect the specific module learning outcomes being assessed. Focus on what students should know, understand, or be able to do as a result of the learning experience.
- *Be clear and concise.* Criteria should be stated simply and unambiguously, for example 'knowledge and understanding', 'application of research skills', or 'use of evidence in argument'. If certain elements such as creativity, clarity of argument, or use of writing conventions (e.g., spelling, punctuation, grammar) are integral to the success of the task and its intended learning outcomes, they should be explicitly included as criteria.
- *One idea per criterion.* Each criterion should address a single skill, knowledge area, or attribute. Avoid combining multiple aspects (e.g., don't say 'critical and creative thinking' in one criterion).
- *Avoid quality judgements in the criteria labels.* Do not embed evaluative language (e.g., 'excellent', 'effective', 'well-developed') in the criterion itself. These should be reserved for the performance level descriptors.

- *Unpack complex criteria.* If a criterion seems broad or abstract, unpack it into more specific components or sub-attributes before defining performance standards. Ask: how, specifically, will the students demonstrate this learning outcome?

Step 4. Identify the threshold requirements

This step involves developing performance standards, which are clear descriptions of the different levels of quality that can be demonstrated for each assessment criterion. These standards form the basis for making informed and consistent judgements about student work. Performance standards describe how well the criterion has been achieved, as opposed to indicating whether a student has fully met a criterion. They outline observable and measurable differences in performance and are typically structured along a scale, such as '*excellent*', '*good*', '*satisfactory*', '*needs improvement*'. We recommend educators to:

- *Ensure clear progression across levels.* Each level should clearly distinguish what is expected. Avoid vague or overlapping descriptors.
- *Consistency.* Keep the number of performance levels consistent across criteria
- *Be specific to the context.* Write standards in a way that makes sense for the discipline, task type, and student level.

Step 5. Identify the standard descriptors

The standard descriptors articulate what performance looks like at different standards providing clarity and transparency for assessors and assessees. They should outline the key characteristics of performance at each level clearly, be specific, and be achievable. Their purpose is to guide judgement, support feedback, and help students understand what is expected of them. Therefore, we recommend educators to:

- *Pitch standards at a reasonable level.* Descriptors should be challenging but achievable. Avoid setting the bar so high that few students can reach the top level, or so low that most students achieve the highest rating with ease.
- *Frame standards positively.* Describe what students should do at each level, rather than what they should avoid. This provides clearer direction and supports a growth-oriented learning environment.
- *Be specific but avoid over-precision.* Aim for specificity in describing performance but avoid unnecessary complexity or oversimplifying complex skills.
- *Define observable behaviours and qualities.* Use verbs that capture student actions (e.g., 'explains', 'analyses', 'applies', 'identifies', 'interprets') and qualifiers or modifiers to define the level of achievement (e.g., 'some', 'appropriate', 'consistent', 'limited', 'comprehensive').
- *Reflect higher-level thinking where appropriate.* At higher levels of achievement, descriptors may include additional cognitive demands, such as originality, depth of analysis, or metacognitive awareness (e.g., reflecting on one's own approach or reasoning).

Step 6. Get feedback

Rubric development is an iterative process, and initial versions often require adjustment. Trialling helps identify ambiguities, overlaps, inconsistencies, or unintended biases in criteria or standards. Calibration and moderation activities also enhance assessment reliability and consistency among assessors. Among the reasons to seek feedback and iterate rubrics:

- To ensure clarity and interpretability of descriptors.
- To assess whether the rubric reflects appropriate technical difficulty and is aligned with learning outcomes.
- To confirm that different assessors interpret the rubric similarly.
- To identify whether the rubric provides useful feedback to students.

4.2 Rubric development checklist

This checklist in Table 3.4.4 provides a structured guide for developing rubrics. It is designed to help ensure that all critical elements, such as purpose, criteria, performance levels, descriptors, and validity, are systematically addressed. Users are encouraged to review each item carefully, make necessary adjustments, and consider the rubric as an evolving tool that can be refined based on feedback and assessment outcomes.

Finally, rubrics should be viewed as evolving tools rather than fixed instruments. Regular review, informed by student performance data, assessor feedback, and changing disciplinary expectations, is essential to ensure that rubrics remain valid, inclusive, and educationally effective within engineering programmes.

Table 3.4.4: Rubric development checklist

1.	*Define the purpose*
☐	Have I identified what the rubric will assess (e.g., project, presentation, essay)?
☐	Is the rubric aligned with learning outcomes?
2.	*Determine the type of rubric*
☐	Have I chosen the most suitable format (holistic or analytic)?
☐	Does this format best support the assessment purpose?
3.	*Identify assessment criteria*
☐	Have I clearly identified the key skills, behaviours, or knowledge areas to be assessed?
☐	Are criteria observable and measurable?
4.	*Develop threshold requirements and standards*
☐	Have I defined the number of performance levels (e.g., as shown in step 4 'excellent' to 'needs improvement')?

☐	Are the performance levels clearly described and differentiated?
5.	*Write descriptors for each level*
☐	Do descriptors for each criterion and level clearly explain the expectations?
☐	Are the descriptors specific, consistent in tone, and free from ambiguous language?
6.	*Review for validity and reliability*
☐	Have I ensured the rubric measures what it intends to measure?
☐	Would different assessors interpret and apply it consistently?
7.	*Pilot and revise*
☐	Have I tested the rubric with sample student work or in a trial run?
☐	Have I collected feedback and made necessary revisions?
8.	*Communicate with students*
☐	Have I shared the rubric with students before the assessment?
☐	Have I explained how the rubric will be used to assess their work?
9.	*Reflect and improve*
☐	Have I reviewed the rubric's effectiveness after use?
☐	Have I identified areas for improvement in future iterations?

5 Rubric examples

To illustrate the guidance provided within this chapter, below are examples of a holistic (Tables 3.4.5 and 3.4.6) and analytic rubrics (Tables 3.4.7-3.4.9). The aim of these rubrics is to evaluate the degree to which an assessment rubric supports and promotes skill development for learners being assessed and for educators using the tool to guide instruction and feedback.

Table 3.4.5: Example of a holistic rubric to assess the effectiveness of a rubric

Criterion	*Descriptor*
Highly effective	The rubric is a highly effective assessment tool. It clearly articulates performance criteria and levels with precise, discipline-appropriate language. It aligns directly with learning outcomes and promotes transparency, fairness, and consistency. The levels are distinct and progressive, allowing for nuanced feedback. The rubric supports formative and/or summative assessment, and it has been tested or reviewed for reliability and validity.

Effective	The rubric is a solid, functional tool for assessment. Criteria and levels are clear and generally align with learning outcomes. It enables fair and consistent evaluation with minimal ambiguity. There may be minor issues in clarity or redundancy, but overall, the rubric facilitates meaningful feedback and student understanding of expectations.
Partially effective	The rubric is functional, but its effectiveness is limited. Some criteria may be vague, overlapping, or misaligned with learning outcomes. Descriptors at performance levels may lack clarity or differentiation, making consistent scoring difficult. The rubric may offer limited support for learning and feedback. Revisions are needed to improve usability and fairness.
Ineffective	The rubric is ineffective or confusing as an assessment tool. Criteria are unclear, overly general, or unrelated to learning outcomes. Performance levels are indistinct or missing, making consistent evaluation impossible. The rubric is ambiguous and does not support constructive feedback. Significant redesign is required.

Table 3.4.6: Example of a holistic rubric to assess working in a team (from Rogers, 2010)

Unsatisfactory	*Developing*	*Satisfactory*	*Exemplary*
Does not collect any information that relates to the topic.	Collects very little information; only some relates to the topic.	Collects some basic information; most relates to the topic	Collects a great deal of information; all relates to the topic.
Does not perform any duties of assigned team role.	Performs very little of assigned duties.	Performs nearly all assigned duties.	Performs all duties of assigned team role.
Always relies on others to do the work.	Rarely does the assigned work; often needs reminding.	Usually does the assigned work; rarely needs reminding	Always does the assigned work without having to be reminded.
Is always talking; never allows anyone else to speak.	Usually doing most of the talking; rarely allows others to speak.	Listens, but sometimes talks too much.	Listens and encourages others to participate.

Table 3.4.7: Example of an analytic rubric that varies in weight to assess the effectiveness of a rubric. The numerical ranges shown within each performance level represent the number of points available for that criterion based on its weighting. For example, "Alignment with learning outcomes" is weighted at 25%, contributing a maximum of 25 points to the overall score of 100 points.

Criteria	*Inadequate (0 - 49%)*	*Good (50 - 69%)*	*Excellent (70 - 79%)*	*Outstanding (80 - 100%)*
Alignment with learning outcomes (25%)	*1 – 12 pts*	*13 – 15 pts*	*16 – 17 pts*	*18 – 25 pts*
	Criteria are poorly aligned or unrelated to learning outcomes.	Criteria are marginally aligned but may be overly general or inconsistent.	Mostly aligned with learning outcomes; minor disconnects in criteria.	Clearly and directly aligned with develop-mental learning outcomes; each criterion reflects targeted skills & competencies.
Clarity and specificity of language (20%)	*0 – 9 pts*	*10 – 13 pts*	*14 – 15 pts*	*16 – 20 pts*
	Descriptors are unclear, inconsistent, or overly specific.	Language is sometimes vague or inconsistent; may confuse learners.	Mostly clear language; some descriptors could be refined.	Language is precise, learner-friendly, and clear across all levels.
Progression of skill development (25%)	*1 – 12 pts*	*13 – 15 pts*	*16 – 17 pts*	*18 – 25 pts*
	No clear progression; levels appear arbitrary or repetitive.	Developmental progression is weak or uneven.	Levels reflect development, but the descriptors distinguishing one level from the next are too similar.	Levels represent a meaningful and realistic developmental progression of skill mastery.

Criteria	*Inadequate (0 - 49%)*	*Good (50 - 69%)*	*Excellent (70 - 79%)*	*Outstanding (80 - 100%)*
Usefulness for intended purpose (15%)	*0 – 7 pts*	*8 – 10 pts*	*11 pts*	*12 – 15 pts*
	Provides little or no value for grading or giving specific feedback depending on the intended use.	Allows for general feedback but limits the ability of specific feedback.	Some support for specific feedback but does not enable students to reflect on their own work.	Encourages self-assessment, goal-setting, and constructive feedback.
Usability and practicality (15%)	*0 – 7 pts*	*8 – 10 pts*	*11 pts*	*12 – 15 pts*
	Unwieldy or confusing to apply; inconsistent scoring likely.	Some usability concerns; rubric may be difficult to implement.	Rubric functions well; minor issues with consistency or application.	Easy to apply reliably; promotes consistency across evaluators; adaptable.

Unlike the holistic rubric shown in Table 3.4.6, which provides a single overall judgement of teamworking performance based on a broad description of performance, the rubric in Table 3.4.8 evaluates multiple dimensions of teamworking separately. Each criterion is assigned the same numerical weight and is scored on a four-point scale from 1 (Unsatisfactory) to 4 (Exemplary). The scores awarded for each criterion are then summed to produce a total score.

When used formatively, the primary purpose of this rubric is to provide detailed feedback on specific aspects of teamworking, helping students identify strengths and areas for improvement. If used summatively, the total score must be mapped to grade boundaries or performance categories established by the module team or institution. For example, a rubric containing five criteria would generate a total score ranging from 5 to 20 points, which could then be translated into pass, merit, or distinction classifications. The specific thresholds should be determined during rubric design and communicated clearly to students before the assessment.

Table 3.4.8: Example of an analytic rubric to assess working in a team, using criteria that have the same numerical weight (from Rogers 2010)

	Unsatisfactory 1 pt	*Developing 2 pts*	*Satisfactory 3 pts*	*Exemplary 4 pts*
Research & gather information	Does not collect any information that relates to the topic.	Collects very little information; some relates to the topic.	Collects some basic information; most relates to the topic.	Collects a great deal of information; all relates to the topic.
Fulfil team role duties	Does not perform any duties of assigned team role.	Performs very little duties.	Performs nearly all duties.	Performs all duties of assigned team role.
Share in work of team	Always relies on others to do the work.	Rarely does the assigned work; often needs reminding.	Usually does the assigned work; rarely needs reminding.	Always does the assigned work without having to be reminded.
Listen to other teammates	Is always talking; never allows anyone else to speak	Usually doing most of the talking; rarely allows others to speak.	Listens, but sometimes talks too much.	Listens and speaks a fair amount.

Table 3.4.9: Sample rubrics designed for analysis of an accrediting body criteria for a project design module, using criteria that have the same numerical weight (from Mahajan & Bansal, 2023)

	Unsatisfactory 1 pt	*Developing 2 pts*	*Satisfactory 3 pts*	*Exemplary 4 pts*
Ability to identify a real life/societal problem to be addressed, using engineering skills and define objectives.	Unable to identify problem statement.	Able to identify problem statement but not able to define clear objectives.	Problem statement is identified but objectives are only partially defined.	Problem statement is identified/defined in a precise manner; objectives are complete and well-defined.

	Unsatisfactory 1 pt	*Developing 2 pts*	*Satisfactory 3 pts*	*Exemplary 4 pts*
Ability to produce and enlist a wide set of problem-specific potential design solutions along with project-specific schedule and budget to meet functional requirement of societal and environmental needs.	Unable to produce any possible solution to defined problem. Unable to develop neither schedule nor budget.	Only one possible solution is presented. Technicalities are not clear. Able to develop either schedule or budget.	Alternative solutions are presented but not able to pick optimal design solution with required justification. Developed both schedule and budget but these are unrealistic.	Capable of generating and comparing solutions to identify the optimal design. Able to define hardware and software requirements through systematic literature review. Detailed and realistic schedule and budget are developed.
Ability to create a prototype to solve a real-life problem.	Unable to identify process tools to develop prototype.	Develops a prototype but does not cater to solve real life problems.	Prototype solves 70% of real-life problem.	The developed prototype is capable to solve real-life problem.
Ability to systematically evaluate designs to meet societal and environmental needs.	No prototype development and validation are done.	Prototype is developed but unable to validate it against real-life problems.	Partial validation is done.	Complete module integration, verification, and validation of prototype.

6 Concluding remarks

Assessment rubrics are invaluable tools that bridge the gap between teaching, learning, and evaluation by clearly defining expectations and standards of quality. As explored in this chapter, the development and implementation of effective rubrics require careful consideration of their purpose, design, and impact on both assessees and assessors. Whether used formatively to guide and support student progress or summatively to assign grades, rubrics must be thoughtfully aligned with learning outcomes and constructed to promote fairness, transparency, and consistency.

Holistic and analytic rubrics each offer distinct advantages and challenges, and the choice between them should be informed by the nature of the task, the skills being assessed, and the type of feedback desired. Critical to the success of any rubric is clarity — clear, measurable criteria paired with detailed performance descriptors that articulate expectations at varying levels of achievement. Moreover, involving multiple stakeholders, particularly assessors and assessees, in the rubric development and calibration processes enhances reliability and fosters shared understanding.

Best practices in rubric development emphasise a systematic approach: defining purpose, aligning criteria with learning outcomes, articulating performance standards, and iteratively refining the tool through feedback and trialling. Effective rubrics not only facilitate reliable assessment but also contribute significantly to student learning by providing actionable feedback and encouraging self-reflection.

Ultimately, rubrics serve as a form of educational contract that guides both teaching and learning. When designed and applied thoughtfully, they promote equity, support skill development, and uphold academic standards, thus making them essential instruments in educational practice.

References

Andrade, H. L., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. Practical Assessment, Research & Evaluation, 10(3), 1–11. https://doi.org/10.7275/g367-ye94

Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. Educational Review, 67 (3), 343–368. https://doi.org/10.1080/00131911.2014.929565

Biggs, J., & Tang, C. (2011). Teaching for quality learning at university (4th ed.). McGraw-Hill/Open University Press.

Cockett, A., & Jackson, C. (2018). The use of assessment rubrics to enhance feedback in higher education: An integrative literature review. Nurse Education Today, 69, 8–13. https://doi.org/10.1016/j.nedt.2018.06.022

Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. Assessment & Evaluation in Higher Education, 42(3), 347–360. https://doi.org/10.1080/02602938.2015.1111294

Hansson, E. E., Svensson, P. J., Strandberg, E. L., Troein, M., & Beckman, A. (2014). Inter-rater reliability and agreement of rubrics for assessment of scientific writing. Education, 4 (1), 12–17. https://doi.org/10.5923/j.edu.20140401.03

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

Mahajan, R., & Bansal, D. (2023). Designing performance metrics and rubrics to assess student outcome attainment in engineering project design course. Journal of Education, 203(2), 459–467. https://doi.org/10.1177/00220574211032587

McKenna, S. (2007). Assessment rubrics as a means of making academic expectations explicit. The Journal of Independent Teaching and Learning, 2, 22–30.

Oakleaf, M. (2009). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. Journal of the American Society for Information Science and Technology, 60 (5), 969–983. https://doi.org/10.1002/asi.21030

O'Connell, B., De Lange, P., Freeman, M., Hancock, P., Abraham, A., Howieson, B., & Watty, K. (2016). Does calibration reduce variability in the assessment of accounting learning outcomes? Assessment & Evaluation in Higher Education, 41 (3), 331–349. https://doi.org/10.1080/02602938.2015.1008398

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

Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. Educational Research Review, 30, Article 100329. https://doi.org/10.1016/j.edurev.2020.100329

Popham, W. J. (1997). What's wrong — and what's right — with rubrics. Educational Leadership, 55(2), 72–77.

Rogers, G. (2010). Developing rubrics. ABET. https://provost.fiu.edu/apa/assessment/resources/rubrics-and-curriculum-maps/_assets/rubrics/Developing%20Rubrics%20in%20Engineering%20-ABET.pdf

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

University of Reading. (2024). Guidance on how to design and use rubrics. Centre for quality support and development. https://www.reading.ac.uk/cqsd/-/media/project/functions/cqsd/documents/ade/tandl-resources/prp-assessment-criteria-and-rubrics.pdf

AI statement

During the preparation of this work, the authors used ChatGPT in order to ideate Tables 3.4.5 and 3.4.7. After using this tool, the authors reviewed, edited and made the content their own as needed. They take full responsibility for the content of the publication.

Contributor statement

Conceptualisation	Folashade Akinmolayan Taiwo, Erik Vellinga
Methodology	Folashade Akinmolayan Taiwo, Erik Vellinga
Project administration	Folashade Akinmolayan Taiwo, Erik Vellinga
Visualisation	Folashade Akinmolayan Taiwo, Erik Vellinga
Writing - original draft	Folashade Akinmolayan Taiwo, Erik Vellinga
Writing - reviewing and editing	Folashade Akinmolayan Taiwo, Erik Vellinga