

Designing transparent evaluation criteria and rubrics with AI

Author: Zsolt Lengyel, iTStudy Hungary Training and Research Centre

Reviewer: Mária Hartyányi, iTStudy Hungary Training and Research Centre

Learning objective

Clear criteria show learners what good work looks like before they start, and make the teacher's judgement consistent and explainable afterwards. The aim of this material is to show how AI can support teachers in designing evaluation criteria and rubrics that are observable, based on learning outcomes and understandable to learners. By the end of the material, teachers will be able to derive criteria from learning outcomes, use AI to draft and refine a rubric with performance levels and descriptors, check the draft for alignment, fairness and transparency, and adapt the final version to their subject, their learners' level and the vocational context – while keeping the final responsibility for the criteria themselves.

Introduction

In vocational education, much of assessment concerns learning products and performances: a weld, an installed circuit, a conversation with a customer, a loaf of bread. They are hard to judge fairly if the criteria exist only in the teacher's head. In innovative assessment methods there are two simple rules: the criteria must be planned in advance, and learners must know them as early as the preparation phase. Assessment should also be objective, reliable and valid – independent of the person who assesses, consistent, and measuring what it is meant to measure.

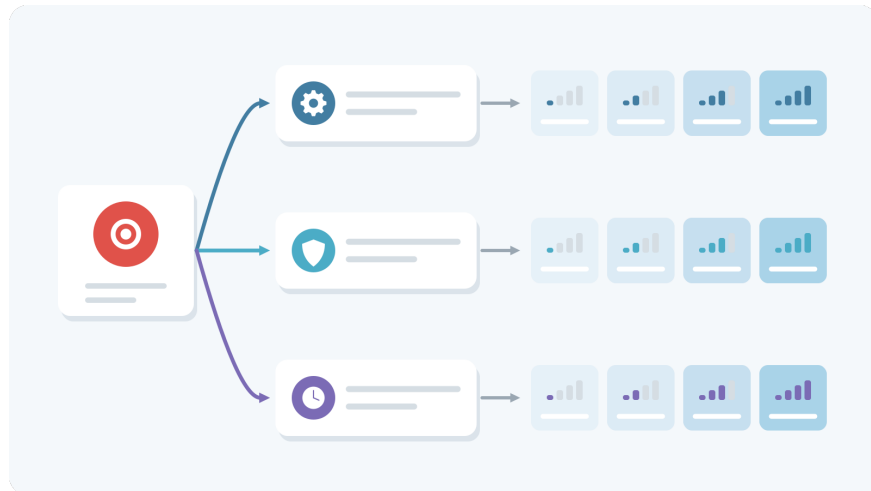
Writing good criteria and descriptors is demanding and time-consuming work, and this is where AI can help. A language model can propose criteria, draft descriptors for several performance levels in minutes, rewrite them in learner-friendly language and point out weaknesses in a draft.

To use a metaphor, AI is an excavator for intellectual work: it multiplies what we can do, but we still have to know where to dig. This material focuses on assessment design and writing feedback on learners' work is the topic of M4-T2-MLC2.

From learning outcomes to criteria

Start from the learning outcome

Criteria are not invented from scratch: they are derived from learning outcomes, which state what the learner knows, understands and is able to do after a learning process, in terms of knowledge, skills, and responsibility and autonomy (EQF). If the learning outcome is "The learner installs a two-way lighting circuit according to the wiring diagram and the safety regulations", the criteria must show how this becomes visible in the product or the performance: correct connections, safe working practice, neat cable routing, testing before switching on, and documentation.



Each learning outcome is broken down into observable criteria, and each criterion into performance levels

What makes a good criterion

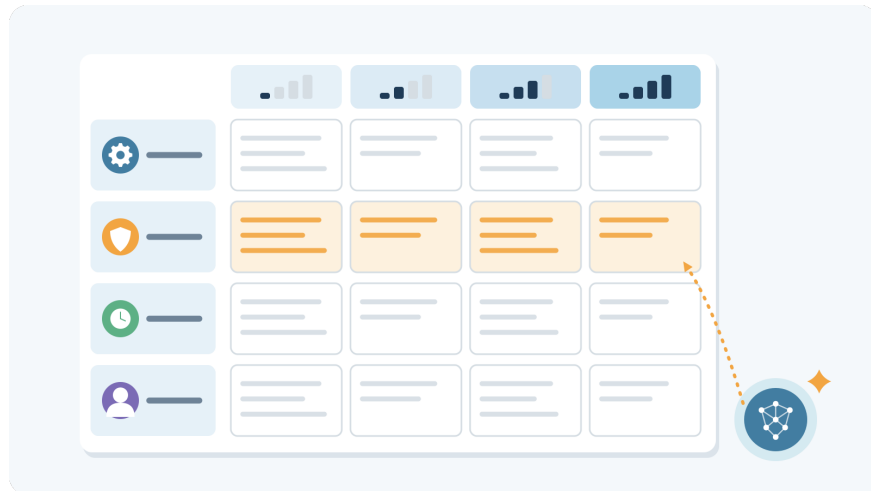
- **Aligned:** it follows from a learning outcome; nothing is assessed that was not taught or intended.
- **Observable:** it describes something that can be seen, measured or heard in the product or performance, not an inner state of the student (e.g. “shows interest”).
- **Distinct:** criteria do not overlap, so the same mistake is not penalised twice.
- **Complete:** together, the criteria cover the essential elements of the outcome, including safety and professional standards.
- **Understandable:** a learner at the given level understands what is expected.

These requirements follow Brookhart’s (2013) guidance on rubrics, and they also serve objectivity, reliability and validity.

Drafting a rubric with AI

Choosing the format

A **checklist** records whether something is present (yes/no); it suits safety checks and procedural steps. An **analytic rubric (assessment matrix)** assesses each criterion separately on several performance levels and gives the most information for learning.



An analytic rubric: criteria in rows, performance levels in columns and a descriptor in every cell
– AI can draft the descriptors row by row

A **holistic rubric** describes overall quality on a single scale; it is quicker to use but offers less guidance. (e.g., rating an essay from 1 to 5 based on a single description).

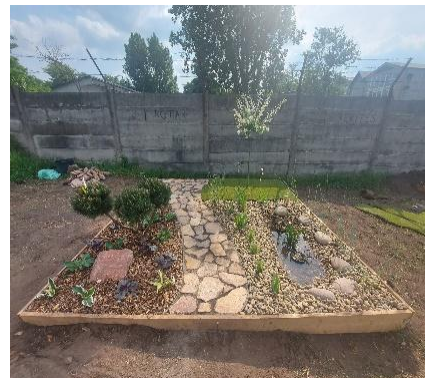
A **single-point rubric** describes only the expected level and leaves space for comments on what falls short of it or goes beyond it – **a practical format for formative assessment.**

Example

In a park construction project the plans were evaluated by the mentor who used a single-point method, with comments and advice written for the plan helping further work.



Design



Implementation

Source: Sára Ekert, agricultural engineer teacher

The construction was evaluated by a professional committee (mentor + two specialist teachers) according to the analytic rubric including the evaluation criteria such as the accuracy of the dimensions, compliance with the plan, the safety of the equipment, etc. at two levels (outstandingly performed/performed).

A prompt that gives the AI what it needs

The five-step framework – task, context, references, evaluation, iteration – works well here. Tell the AI **what** to create and **for whom** (learners' age, year

and prior knowledge), give it **references** (the learning outcomes, the exam requirements, your task description, an earlier rubric), then **evaluate** the output critically and **iterate**. Adding one good and one weak example of a descriptor improves precision considerably.

Example prompt for a rubric draft

"You are an experienced vocational teacher of electricians. Create an analytic rubric for a practical task: second-year apprentices (aged 16–17) install a two-way lighting circuit on a training board. Learning outcomes: [paste]. Criteria: correctness of connections, safe working practice, cable routing and workmanship, testing and documentation. Use four performance levels. Write descriptors that describe observable features of the work, use parallel wording across the levels and are understandable to the apprentices. Mark any criterion that must be met in full regardless of the other scores (e.g. safety). Present the rubric as a table. Then list three weaknesses of your own draft."

Performance levels and descriptors

Three to five levels are typical; an even number (for example four) removes the tempting middle option. Good descriptors describe the quality of the work – what can be seen at that level – instead of using only evaluative words ("good", "weak") or mere counting ("three mistakes"). The wording should change in parallel from level to level, so that learners can see exactly what distinguishes one level from the next. An example for one criterion of a welding practice piece (appearance and uniformity of the weld bead):

- **Level 1 – not yet sufficient:** the width and height of the bead vary strongly; undercut, porosity or lack of fusion is visible; the piece does not meet the requirements.
- **Level 2 – developing:** the bead is mostly continuous, but its width varies noticeably; there are minor surface defects at the start and end of the weld.
- **Level 3 – meets the standard:** the bead is continuous and even in width and height; the start and end are properly filled; no surface defects are visible.
- **Level 4 – exceeds the standard:** as level 3, and the bead is uniform along its whole length, matching the reference sample; the learner can explain how they controlled travel speed and torch or electrode angle.

AI drafts such rows quickly, but only the teacher can judge whether "even in width" is precise enough for the given process and material, or whether the workshop's reference samples and the relevant standard should be named.

Checking alignment, fairness and transparency

The first draft is only a starting point. AI is also useful as a critical reader – if we ask neutral rather than leading questions. "*Isn't this a good rubric?*" invites

agreement; “*What are the weaknesses of this rubric?*” invites analysis.



Checking the draft: alignment with the learning outcomes, clarity, fairness and consistency between assessors

Useful checks:

- **Aligned with the curriculum**
“Map each criterion to the learning outcomes. Which outcome is not assessed? Which criterion does not belong to any outcome?”
- **Clarity**
“Which words are vague or open to interpretation? Rewrite the descriptors for a 16-year-old apprentice.”
- **Fairness**
“Does any criterion reward something that is not part of the outcome – for example spelling in a practical task? Does the wording disadvantage learners with weaker reading skills or a different first language?”
- **Weighting**
“Do the points reflect the importance of the criteria? Is safety treated as a condition that must be met?”
- **Consistency**
Try the rubric on two or three anonymised or AI-invented sample products of different quality; ideally, two colleagues score them independently – where they disagree, the descriptor needs more work.

Finally, transparency means that learners see and understand the criteria before they start. AI can turn the rubric into a learner-friendly version (“I can...” statements) or a self-assessment checklist; it is even better to discuss the criteria with the learners, who can then use them for self- and peer assessment.

The teacher’s responsibility

AI does not know your learners, your workshop, the exact exam requirements or the expectations of local employers – unless you provide them, and even then it may misinterpret them. Its descriptors can sound convincing while being technically wrong or unrealistic. The teacher therefore decides:

- whether the criteria are professionally correct and suit the subject and the vocational context;

- whether the levels are realistic for the learners' year and prior knowledge;
- how points or levels are converted into a grade, and which criteria are non-negotiable;
- when the rubric is revised – for example, based on the experience of its first use.



AI drafts, the teacher decides: the approved criteria are shared with learners before the task

AI support for rubrics is increasingly built into learning platforms: in Moodle, rubrics are an advanced grading method, and Gemini in Google Classroom can draft a rubric from the assignment description where the school has enabled it. Whatever the tool, designing criteria requires no personal data – the task description and the learning outcomes are enough. Designing criteria with AI is, however, different from evaluating learners with AI: the EU AI Act classifies AI systems intended to evaluate learning outcomes as high-risk. The criteria, and the judgement based on them, remain the teacher's.

Takeaways for teachers

- Start from the learning outcomes: every criterion should be observable and traceable to an outcome.
- Give the AI context and references: learners' level, task, learning outcomes, exam requirements, examples.
- Ask for descriptors that describe quality with parallel wording across levels – and check their professional correctness.
- Use AI as a critical reader with neutral questions: alignment, clarity, fairness, weighting.
- Share the criteria with learners before the task, and keep the final responsibility for them.

References

- Brookhart, S. M. (2013): How to Create and Use Rubrics for Formative Assessment and Grading. ASCD.
- Farkas, É. (2018): Mérés-értékelés kézikönyv. Tanulási eredmények mérése és értékelése a szakképzési mobilitási gyakorlatokban [Assessment handbook]. Tempus Public Foundation.
- Council Recommendation of 22 May 2017 on the European Qualifications Framework for lifelong learning (2017/C 189/03).

- Redecker, C. (2017): European Framework for the Digital Competence of Educators: DigCompEdu. Publications Office of the European Union (area 4.1: Assessment strategies).
- Regulation (EU) 2024/1689 (Artificial Intelligence Act), Annex III, point 3.
- iTStudy Hungary: Innovatív értékelési módszerek [Innovative assessment methods], course module.
- Lengyel, Zs. (2026): Concrete applications for content generation. VITA training presentation.

Links

- Moodle Docs – Rubrics: docs.moodle.org/en/Rubrics
- Gemini in Google Classroom: support.google.com/edu/classroom
- DigCompEdu: joint-research-centre.ec.europa.eu/digcompedu_en

Copyright © 2026 VITA Consortium

This work may be reused and adapted under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).



The illustrations were generated by Anthropic Claude.

Funded by the European Union. Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the National Agency.