Diseño de criterios de evaluación y rúbricas transparentes con IA

Autor: Zsolt Lengyel, iTStudy Hungary Training and Research Centre

Revisora: Mária Hartyányi, iTStudy Hungary Training and Research Centre

Objetivo de aprendizaje

Unos criterios claros muestran al alumnado cómo es un buen trabajo antes de empezar y, después, hacen coherente y explicable el juicio del docente. El objetivo de este material es mostrar cómo la IA puede ayudar a los docentes a diseñar criterios de evaluación y rúbricas que sean observables, se basen en los resultados de aprendizaje y resulten comprensibles para el alumnado. Al terminar el material, los docentes serán capaces de derivar criterios a partir de los resultados de aprendizaje, utilizar la IA para redactar y perfeccionar una rúbrica con niveles de desempeño y descriptores, revisar el borrador en cuanto a alineación, equidad y transparencia, y adaptar la versión final a su materia, al nivel de su alumnado y al contexto profesional, conservando la responsabilidad última sobre los propios criterios.

Introducción

En la formación profesional, gran parte de la evaluación se centra en productos y desempeños: una soldadura, un circuito instalado, una conversación con un cliente, una hogaza de pan. Es difícil evaluarlos con equidad si los criterios solo existen en la cabeza del docente. Los métodos de evaluación innovadores se basan en dos reglas sencillas: los criterios deben planificarse de antemano y el alumnado debe conocerlos ya desde la fase de preparación. Además, la evaluación debe ser objetiva, fiable y válida: independiente de la persona que evalúa, coherente y capaz de medir lo que pretende medir.

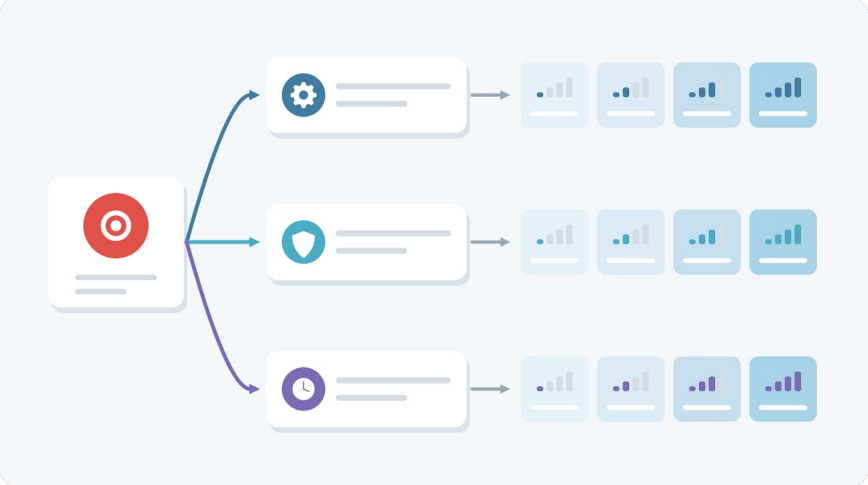
Redactar buenos criterios y descriptores es un trabajo exigente y laborioso, y aquí es donde la IA puede ayudar. Un modelo de lenguaje puede proponer criterios, redactar en pocos minutos descripciones para varios niveles de desempeño, reescribirlas en un lenguaje accesible para el alumnado y señalar los puntos débiles de un borrador.

(La redacción de retroalimentación sobre los trabajos del alumnado es el tema de M4-T2-MLC2).

De los resultados de aprendizaje a los criterios

Parta del resultado de aprendizaje

Los criterios no se inventan de la nada: se derivan de los resultados de aprendizaje, que expresan lo que el estudiante sabe, comprende y es capaz de hacer al término de un proceso de aprendizaje, en términos de conocimientos, destrezas y responsabilidad y autonomía (MEC). Si el resultado de aprendizaje es “El estudiante instala un circuito de alumbrado conmutado conforme al esquema de conexiones y a las normas de seguridad”, los criterios deben mostrar cómo se hace visible en el producto o en el desempeño: conexiones correctas, trabajo seguro, tendido ordenado del cableado, comprobación antes de la puesta en tensión y documentación.



Cada resultado de aprendizaje se desglosa en criterios observables, y cada criterio, en niveles de desempeño

Qué hace que un criterio sea bueno

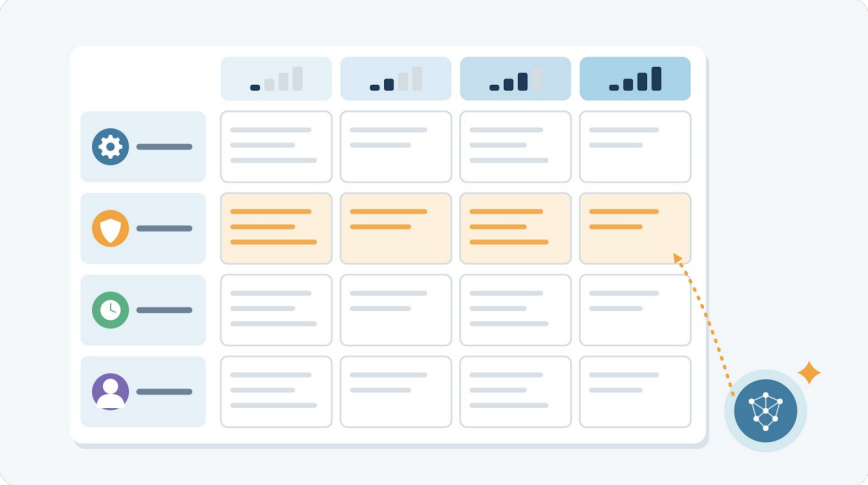
- **Alineado:** se deriva de un resultado de aprendizaje; no se evalúa nada que no se haya enseñado o previsto.
- **Observable:** describe algo que se puede ver, medir u oír en el producto o en el desempeño, no un estado interior del estudiante (p. ej., "muestra interés").
- **Diferenciado:** los criterios no se solapan, de modo que un mismo error no se penaliza dos veces.
- **Completo:** los criterios abarcan los elementos esenciales del resultado, incluidas la seguridad y las normas profesionales.
- **Comprensible:** un estudiante del nivel correspondiente entiende qué se espera de él.

Estos requisitos siguen las orientaciones de Brookhart (2013) sobre las rúbricas y, además, contribuyen a la objetividad, la fiabilidad y la validez.

Cómo redactar una rúbrica con IA

Elección del formato

Una **lista de cotejo** registra si algo está presente o no (sí/no); es adecuada para comprobaciones de seguridad y pasos de un procedimiento. Una **rúbrica analítica (matriz de evaluación)** evalúa cada criterio por separado en varios niveles de desempeño y es la que más información aporta para el aprendizaje.



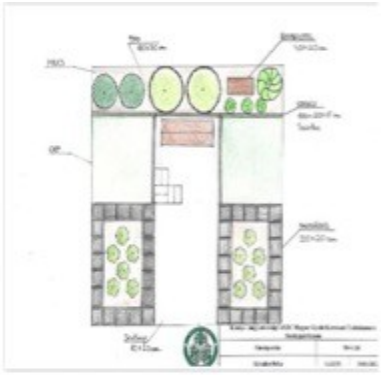
Una rúbrica analítica: criterios en las filas, niveles de desempeño en las columnas y un descriptor en cada celda; la IA puede redactar los descriptores fila a fila

Una **rúbrica holística** describe la calidad global en una única escala; es más rápida de aplicar, pero ofrece menos orientación (p. ej., calificar una redacción del 1 al 5 a partir de una única descripción).

Una **rúbrica de punto único** describe solo el nivel esperado y deja espacio para comentar lo que no llega a ese nivel o lo supera: **un formato práctico para la evaluación formativa.**

Ejemplo

En un proyecto de construcción de un parque, el mentor evaluó los planos mediante una rúbrica de punto único, con comentarios y consejos escritos sobre el plano que ayudaron a continuar el trabajo.





Diseño Ejecución

Fuente: Sára Ekert, ingeniera agrónoma y docente

La obra fue evaluada por un tribunal profesional (el mentor y dos docentes especialistas) según una rúbrica analítica que incluía criterios de evaluación como la precisión de las dimensiones, la conformidad con el plano o la seguridad del equipamiento, en dos niveles (realizado de forma sobresaliente / realizado).

Un prompt que dé a la IA lo que necesita

El marco de cinco pasos – tarea, contexto, referencias, evaluación, iteración – funciona bien en este caso. Indique a la IA **qué** debe crear y **para quién** (edad,

curso y conocimientos previos del alumnado), proporciónele **referencias** (los resultados de aprendizaje, los requisitos del examen, la descripción de su tarea, una rúbrica anterior) y, a continuación, **evalúe** críticamente el resultado e **itere**. Añadir un ejemplo de descriptor bueno y otro deficiente mejora considerablemente la precisión.

Ejemplo de prompt para un borrador de rúbrica
"Eres un docente de FP con experiencia en la formación de electricistas. Crea una rúbrica analítica para una tarea práctica: aprendices de segundo curso (de 16 a 17 años) instalan un circuito de alumbrado conmutado en un panel de prácticas. Resultados de aprendizaje: [pegar aquí]. Criterios: corrección de las conexiones, trabajo seguro, tendido del cableado y acabado, comprobación y documentación. Utiliza cuatro niveles de desempeño. Redacta descriptores que describan características observables del trabajo, con una formulación paralela entre los niveles, que resulten comprensibles para los aprendices. Señala cualquier criterio que deba cumplirse en su totalidad con independencia de las demás puntuaciones (p. ej., la seguridad). Presenta la rúbrica en forma de tabla. Después, enumera tres puntos débiles de tu propio borrador".

Niveles de desempeño y descriptores

Lo habitual son entre tres y cinco niveles; un número par (por ejemplo, cuatro) elimina la tentadora opción intermedia. Los buenos descriptores describen la calidad del trabajo – lo que se observa en ese nivel – en lugar de limitarse a palabras valorativas ("bueno", "flojo") o a simples recuentos ("tres errores"). La formulación debe cambiar de forma paralela de un nivel a otro, para que el alumnado vea exactamente qué distingue un nivel del siguiente. Un ejemplo para un criterio de una pieza de prácticas de soldadura (aspecto y uniformidad del cordón de soldadura):

- **Nivel 1 – todavía insuficiente:** la anchura y la altura del cordón varían mucho; se observan mordeduras, porosidad o falta de fusión; la pieza no cumple los requisitos.
- **Nivel 2 – en desarrollo:** el cordón es mayoritariamente continuo, pero su anchura varía de forma apreciable; hay pequeños defectos superficiales al inicio y al final de la soldadura.
- **Nivel 3 – cumple el estándar:** el cordón es continuo y uniforme en anchura y altura; el inicio y el final están correctamente rellenos; no se observan defectos superficiales.
- **Nivel 4 – supera el estándar:** como en el nivel 3, y además el cordón es uniforme en toda su longitud y coincide con la muestra de referencia; el estudiante sabe explicar cómo ha controlado la velocidad de avance y el ángulo de la antorcha o del electrodo.

La IA redacta estas filas rápidamente, pero solo el docente puede juzgar si "uniforme en anchura" es lo bastante preciso para el proceso y el material concretos, o si conviene mencionar las muestras de referencia del taller y la norma aplicable.

Revisión de la alineación, la equidad y la transparencia

El primer borrador es solo un punto de partida. La IA también resulta útil como lectora crítica, siempre que le hagamos preguntas neutras y no tendenciosas. "*¿Verdad que es una buena rúbrica?*" invita a darnos la razón; "*¿Cuáles son los puntos débiles de esta rúbrica?*" invita al análisis.



Revisión del borrador: alineación con los resultados de aprendizaje, claridad, equidad y coherencia entre evaluadores

Comprobaciones útiles:

- **Alineación con el currículo**

  "Relaciona cada criterio con los resultados de aprendizaje. ¿Qué resultado no se evalúa? ¿Qué criterio no corresponde a ningún resultado?"

- **Claridad**

  "¿Qué palabras son vagas o se prestan a interpretación? Reescribe los descriptores para un aprendiz de 16 años".

- **Equidad**

  "¿Hay algún criterio que premie algo que no forma parte del resultado, por ejemplo, la ortografía en una tarea práctica? ¿Perjudica la redacción al alumnado con menor competencia lectora o con otra lengua materna?"

- **Ponderación**

  "¿Reflejan los puntos la importancia de los criterios? ¿Se trata la seguridad como una condición de obligado cumplimiento?"

- **Coherencia**

  Pruebe la rúbrica con dos o tres productos de muestra anonimizados o inventados por la IA, de distinta calidad; lo ideal es que dos colegas los puntúen de forma independiente: allí donde discrepen, el descriptor necesita más trabajo.

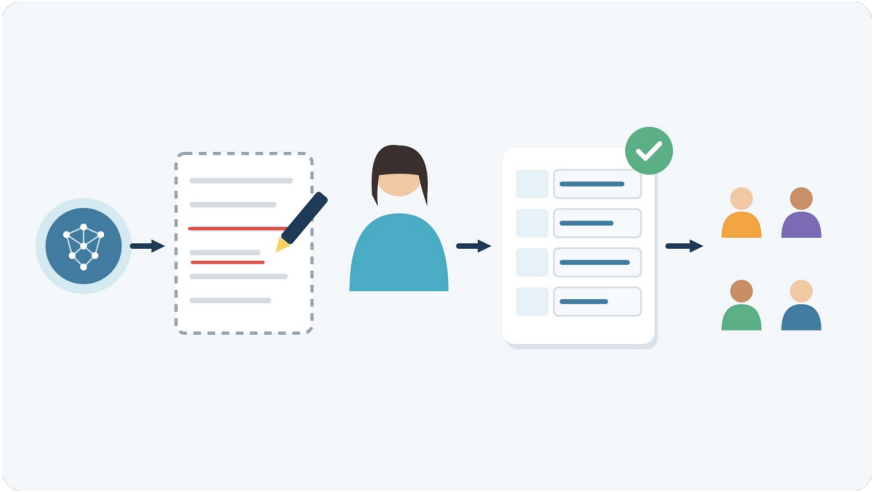
Por último, transparencia significa que el alumnado ve y entiende los criterios antes de empezar. La IA puede convertir la rúbrica en una versión adaptada al

alumnado (afirmaciones del tipo “Soy capaz de…”) o en una lista de autoevaluación; aún mejor es comentar los criterios con el alumnado, que así podrá utilizarlos para la autoevaluación y la evaluación entre iguales.

La responsabilidad del docente

La IA no conoce a su alumnado, su taller, los requisitos exactos del examen ni las expectativas de las empresas de la zona, salvo que usted se lo proporcione, e incluso entonces puede malinterpretarlo. Sus descriptores pueden sonar convincentes y, sin embargo, ser técnicamente erróneos o poco realistas. Por eso es el docente quien decide:

- si los criterios son correctos desde el punto de vista profesional y se ajustan a la materia y al contexto profesional;
- si los niveles son realistas para el curso y los conocimientos previos del alumnado;
- cómo se convierten los puntos o niveles en una calificación y qué criterios son innegociables;
- cuándo se revisa la rúbrica, por ejemplo, a partir de la experiencia de su primer uso.



La IA redacta y el docente decide: los criterios aprobados se comparten con el alumnado antes de la tarea

La ayuda de la IA para crear rúbricas se integra cada vez más en las plataformas de aprendizaje: en Moodle, las rúbricas son un método de calificación avanzado, y Gemini en Google Classroom puede redactar una rúbrica a partir de la descripción de la tarea si el centro lo ha habilitado. Sea cual sea la herramienta, diseñar criterios no requiere datos personales: basta con la descripción de la tarea y los resultados de aprendizaje. Ahora bien, diseñar criterios con IA no es lo mismo que evaluar al alumnado con ella: el Reglamento de Inteligencia Artificial de la UE clasifica como de alto riesgo los sistemas de IA destinados a evaluar resultados de aprendizaje. Los criterios, y el juicio basado en ellos, siguen siendo competencia del docente.

Ideas clave para docentes

- Parta de los resultados de aprendizaje: cada criterio debe ser observable y poder vincularse a un resultado.
- Proporcione a la IA contexto y referencias: nivel del alumnado, tarea, resultados de aprendizaje, requisitos del examen, ejemplos.
- Pida descriptores que describan la calidad con una formulación paralela entre niveles y compruebe que sean correctos desde el punto de vista profesional.
- Utilice la IA como lectora crítica con preguntas neutras: alineación, claridad, equidad, ponderación.
- Comparta los criterios con el alumnado antes de la tarea y conserve la responsabilidad última sobre ellos.

Referencias

- 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 [Manual de evaluación]. Tempus Public Foundation.
- Recomendación del Consejo, de 22 de mayo de 2017, relativa al Marco Europeo de Cualificaciones para el aprendizaje permanente (2017/C 189/03).
- Redecker, C. (2017): European Framework for the Digital Competence of Educators: DigCompEdu. Publications Office of the European Union (área 4.1: Estrategias de evaluación).
- Reglamento (UE) 2024/1689 (Reglamento de Inteligencia Artificial), anexo III, punto 3.
- iTStudy Hungary: Innovatív értékelési módszerek [Métodos de evaluación innovadores], módulo de curso.
- Lengyel, Zs. (2026): Concrete applications for content generation. Presentación de la formación VITA.

Enlaces

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

Derechos de autor © 2026 Consorcio VITA

Esta obra puede compartirse y adaptarse bajo los términos de la Licencia Internacional Creative Commons Atribución/Reconocimiento-CompartirIgual 4.0 (CC BY-SA 4.0).



Las ilustraciones han sido generadas con Anthropic Claude.

Financiado por la Unión Europea. Las opiniones expresadas son únicamente las de los autores y no reflejan necesariamente las de la Unión Europea ni las de la Agencia Nacional.