

3. Creación de instrumentos de evaluación

La evaluación educativa no consiste solo en poner una nota, sino en recoger evidencias, interpretarlas y tomar decisiones pedagógicas con fundamento. En la era de la IA, esta idea cobra todavía más importancia, porque las herramientas digitales pueden ayudar a diseñar instrumentos más precisos, pero también pueden generar una falsa sensación de objetividad si no hay una mirada docente crítica. Por eso, crear instrumentos de evaluación sigue siendo una tarea profundamente profesional, ligada al currículo, a la competencia y al contexto real del aula.

Los instrumentos de evaluación son las herramientas que permiten recoger evidencias de aprendizaje de forma sistemática. En un contexto con IA, su valor aumenta porque ayudan a hacer explícitos los criterios, reducir la subjetividad y generar evidencias más comparables. La IA puede apoyar la redacción, la mejora del lenguaje, la revisión de coherencia o la generación de borradores, pero la decisión final sobre criterios, niveles y ponderaciones debe seguir siendo docente. En evaluación, la clave no es que la IA “corrija”, sino que ayude a diseñar instrumentos más claros, válidos y alineados con los objetivos de aprendizaje.

Por tanto, desde una perspectiva didáctica, un buen instrumento de evaluación debe cumplir cuatro condiciones básicas: estar **alineado con lo que se quiere evaluar**, ser **comprensible para quien lo usa**, permitir **recoger evidencias observables** y **facilitar decisiones coherentes**. La IA puede acelerar la redacción o proponer borradores, pero no sustituye el juicio pedagógico ni la selección intencional de los indicadores. En otras palabras, la tecnología puede ayudar a redactar mejor; no puede decidir por sí sola qué significa aprender bien.

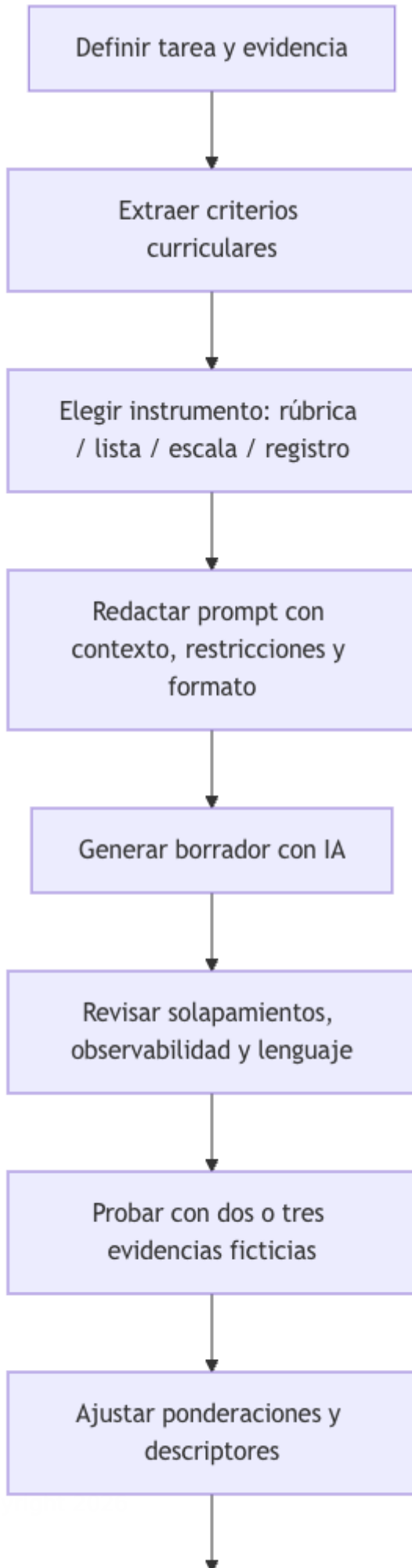
- [3.1. Consejos para crear buenos prompts](#)
- [3.2. Rúbricas](#)
- [3.3. Listas de cotejo](#)
- [3.4. Escalas de valoración](#)



- [3.5. Registros de observación](#)

3.1. Consejos para crear buenos prompts

El mejor uso inicial de la IA en evaluación no es la corrección, sino la **creación de instrumentos**. El **orden** correcto es este: **primero se fijan competencias específicas, criterios de evaluación, evidencias y nivel esperado; después se usa la IA para convertir eso en formatos operativos**. Si se trabaja al revés —pidiéndole a un chatbot “hazme una rúbrica” sin criterio previo— lo habitual es obtener textos aparentemente elegantes, pero genéricos, redundantes o mal alineados con lo que realmente se pretende valorar. Suelen ser buenos borradores iniciales, pero requieren revisión y adaptación docente sustantiva.





Aránzazu Cortés + Chat GPT

La fase crítica no es la generación, sino la **prueba con evidencias ficticias o modelos de respuesta**, porque ahí se detecta si la rúbrica distingue de verdad entre niveles. El uso de few-shot y salidas estructuradas mejora mucho esa iteración.

Por lo tanto, cuando pedimos a una IA que genere un instrumento de evaluación, el resultado depende en gran medida de la calidad del prompt. Un buen prompt no debe limitarse a “hazme una rúbrica”, sino que debe contextualizar la tarea, el nivel educativo, la materia, el tipo de evidencia y el uso que se dará al instrumento. Cuanta más información útil reciba la IA, más ajustada será la propuesta.

Conviene incorporar cinco elementos: **contexto didáctico** (etapa educativa, materia, nivel del alumnado), **criterios o competencias, tipo de evidencia a valorar, formato del instrumento y restricciones**. También es útil indicar el **tono** si el instrumento será **para alumnado, profesorado o ambos**, porque el lenguaje cambia mucho según el destinatario. Si el docente no acota, la IA tiende a generar instrumentos genéricos, correctos en apariencia pero poco útiles para una situación concreta.

Cuanto más precisa sea la instrucción, más útil será la propuesta de la IA. También conviene pedir siempre que el resultado sea revisable y que no invente criterios fuera del currículo.

La fórmula más fiable para profesorado suele ser esta:

Rol + tarea + currículo + evidencia + restricciones + formato + control de calidad.

Ejemplo de patrón:

“ Actúa como especialista en evaluación educativa. Diseña una rúbrica analítica para 2.º ESO, Geografía e Historia, para evaluar un podcast de 3 minutos. Usa 4 criterios alineados con [criterios del currículo], 4 niveles de desempeño, lenguaje comprensible para alumnado y profesorado, descriptores observables y no redundantes. No inventes criterios ajenos al currículo. Devuelve una tabla y un JSON.

Estructura útil de prompt

- Contexto: curso, materia y producto final.
- Objetivo de evaluación: qué se quiere valorar.
- Criterios o competencias.



- Formato deseado: rúbrica, lista de cotejo, escala, registro.
- Nivel de detalle.
- Restricciones: lenguaje simple, sin datos personales, alineado con criterios curriculares.

Ejemplos de prompts bien formulados:

“ Crea una rúbrica analítica para 3.º ESO de Lengua Castellana para evaluar una exposición oral de 3 minutos. Incluye 4 criterios, 4 niveles de desempeño, lenguaje claro para alumnado, y relación con comunicación oral, organización de ideas y uso de recursos expresivos.

“ Genera una rúbrica analítica para 4.º de ESO de Biología para evaluar un informe de laboratorio. Debe incluir 4 criterios, 4 niveles de logro, lenguaje claro para alumnado, y centrarse en presentación de datos, interpretación, vocabulario científico y conclusión.

“ Crea una lista de cotejo para Primaria, 5.º curso, para valorar una exposición oral sobre animales vertebrados. Debe ser breve, observable y útil para observación rápida en clase.

La ventaja de este enfoque es que la IA no improvisa, sino que trabaja dentro de un marco pedagógico definido por el docente. Esto favorece la coherencia curricular y evita instrumentos demasiado abstractos o desalineados con la tarea real.

La IA en educación debe usarse con supervisión docente, con protección de datos y dentro de un enfoque ético y pedagógico. Además, en el ámbito de la evaluación educativa los sistemas de IA pueden ser especialmente sensibles cuando afectan a resultados de aprendizaje o decisiones sobre alumnado.

Una regla muy simple y muy técnica a la vez: **identidad, instrucciones, ejemplos y contexto.**

Continuando con la creación de prompts eficaces, además, se recomienda estructurarlos con secciones claras, usar encabezados o etiquetas, incluir ejemplos diversos de entrada/salida y



añadir el contexto relevante al final; además, cuando el resultado tiene que reutilizarse, conviene pedir la salida en formato estructurado, por ejemplo JSON o tabla.

Un prompt eficaz no dice solo “créame una rúbrica”, sino algo como:

“ Actúa como especialista en evaluación competencial de 2.º de ESO. A partir de este criterio, genera una rúbrica analítica de cuatro niveles, con indicadores observables, lenguaje comprensible para alumnado y familias, y una versión final en tabla y otra en JSON.

Cuando además se añaden dos o tres ejemplos buenos y malos de producto del alumnado, la IA suele discriminar mucho mejor entre niveles de desempeño. La documentación oficial sobre prompting denomina a esto **few-shot learning** y recomienda que los ejemplos cubran una gama diversa de situaciones para que el patrón salga más robusto.

Para que estos instrumentos salgan bien con IA, hay varias reglas prácticas que merece la pena enseñar. La primera es **un criterio, una dimensión**: si un descriptor mezcla “contenido, expresión y creatividad”, la IA tenderá a producir niveles ambiguos y el profesorado tendrá más desacuerdo al corregir. La segunda es **lenguaje observable**: mejor “justifica su respuesta con dos evidencias del texto” que “demuestra comprensión profunda”. La tercera es **anclaje por evidencias**: pedir expresamente que cada nivel incluya señales observables o errores típicos. La cuarta es **portabilidad**: solicitar a la IA versión larga para el profesorado, versión breve para el alumnado y versión tabular o JSON para reutilizarla en hojas de cálculo, formularios o LMS. La recomendación de pedir resultados estructurados está alineada con la práctica oficial de “Structured Outputs”.

¿Qué instrumento conviene en cada caso?



Instrumento	Cuándo usarlo	Qué aporta la IA	Riesgo típico
Rúbrica analítica	Proyectos, escritos, exposiciones, tareas complejas	Redacción de criterios, niveles, versiones para alumnado y profesorado	Descriptores vagos o solapados
Lista de cotejo	Requisitos mínimos, procedimientos, entregables	Convertir criterios en ítems observables tipo sí/no	Reducir calidad a mera presencia
Escala de valoración	Participación, cooperación, hábitos, frecuencia/intensidad	Afinar indicadores graduales y anclas verbales	Ambigüedad en "a veces", "frecuente", etc.
Registro de observación	Procesos, oralidad, laboratorio, trabajo cooperativo	Clasificar observaciones y resumir patrones	Convertir anécdotas en diagnósticos excesivos

La recomendación práctica es usar **rúbricas** para tareas complejas, **listas de cotejo** para control mínimo de calidad, **escalas** para procesos graduales y **registros de observación** para evidencias no capturadas en productos finales. Esa variedad encaja mejor con un enfoque competencial y evita sobrecargar una sola herramienta.

3.2. Rúbricas

La rúbrica es probablemente el instrumento más completo para evaluar tareas complejas. Su principal virtud es que hace visibles los criterios de calidad y describe distintos niveles de desempeño, lo que reduce la arbitrariedad y mejora la transparencia. Es especialmente útil en tareas competenciales, proyectos, presentaciones orales, textos escritos, trabajos cooperativos y producciones creativas.

Teóricamente, la rúbrica tiene un valor importante porque convierte la evaluación en un proceso más formativo. El alumnado no solo recibe una calificación final, sino que puede entender qué se espera, en qué nivel se encuentra y qué debe mejorar. Además, una rúbrica bien diseñada ayuda al profesorado a unificar criterios entre grupos o entre docentes que corrigen una misma tarea.

Con IA, la rúbrica puede:

- generarse como borrador,
- revisarse para mejorar su lenguaje,
- adaptarse a distintos niveles o materias,
- generar versiones para alumnado y profesorado.

Sin embargo, hay que vigilar varios problemas frecuentes: descriptores que se solapan, niveles que no son realmente distintos, criterios demasiado vagos o excesiva cantidad de indicadores. Una rúbrica buena es clara, breve, equilibrada y basada en evidencias observables.

Ejemplo práctico: rúbrica para una infografía

Criterio	Nivel 4	Nivel 3	Nivel 2	Nivel 1
Exactitud del contenido	Toda la información es correcta y relevante.	Hay pequeños errores sin afectar al sentido.	Hay varios errores o alguna información poco relevante.	Contenido incorrecto o muy incompleto.
Organización visual	Muy clara, jerarquía visual eficaz.	Bastante clara y ordenada.	Algo confusa o sobrecargada.	Difícil de seguir.
Síntesis	Resume con precisión sin perder ideas clave.	Resume bien con algunas repeticiones.	Exceso de texto o falta de síntesis.	No sintetiza.
Comunicación	Atrae y facilita la comprensión.	Comunica adecuadamente.	Comunica con dificultad.	No facilita la comprensión.

Ejemplo práctico: rúbrica para una exposición oral



Criterio	Nivel 4	Nivel 3	Nivel 2	Nivel 1
Organización de ideas	Presenta una estructura muy clara, con introducción, desarrollo y cierre bien conectados.	La estructura es clara con pequeños saltos de ideas.	La organización es irregular y cuesta seguir el hilo.	No hay estructura reconocible.
Expresión oral	Habla con seguridad, ritmo adecuado y buena pronunciación.	Se expresa con claridad, aunque con alguna vacilación.	Presenta pausas frecuentes o dificultades de pronunciación.	La exposición resulta difícil de seguir.
Contenido	El contenido es completo, riguroso y bien seleccionado.	El contenido es adecuado, aunque algo superficial.	Hay lagunas o información poco precisa.	El contenido es muy incompleto o incorrecto.
Apoyo visual	Usa apoyos útiles, limpios y coherentes con la exposición.	Los apoyos son adecuados aunque mejorables.	Los apoyos distraen o aportan poco.	No utiliza apoyos o son inadecuados.

Ejemplo de uso docente: la IA puede redactar los niveles, pero el profesor decide si “expresión oral” debe valer más que “apoyo visual”, si el criterio de contenido incluye exactitud conceptual y si el lenguaje es adecuado para la edad. Esa validación es esencial para que la rúbrica mida lo que realmente interesa.

De ahí nace la diferencia entre **rúbrica de diseño** y **rúbrica de uso**. La primera la emplea el docente para pensar y consensuar criterios; la segunda se adapta al momento de aplicación. Muchas veces conviene que no sean idénticas. Por ejemplo, el profesorado puede trabajar con una rúbrica analítica detallada de cinco criterios, mientras que al alumnado se le entrega una síntesis visual de tres dimensiones con ejemplos de calidad. Los sistemas de IA para generar rúbricas han resultado útiles precisamente como **andamio de diseño** inicial, no como instrumento final listo para usar sin edición.

Uso con IA: puedes pedirle que genere los descriptores, pero después conviene revisar si los niveles son realmente distinguibles entre sí.

Plantilla de prompt para generar una rúbrica analítica

“ Diseña una rúbrica analítica para [curso/etapa], [materia], para evaluar [producto]. Usa [n] criterios alineados con [criterios curriculares], [n] niveles, descriptores observables y lenguaje claro. No inventes criterios fuera del currículo. Devuelve tabla y JSON.



- “ Actúa como especialista en evaluación competencial en [etapa educativa] y [materia].Objetivo:
 Crear una rúbrica analítica alineada con este criterio de evaluación:
 [pegar criterio]Contexto:
- Situación de aprendizaje: [describir]
 - Producto o evidencia: [texto oral, informe, maqueta, problema resuelto, etc.]
 - Nivel esperado: [curso/etapa]
 - Número de sesiones: [x]
 - Uso final: evaluación formativa y feedback al alumnadoInstrucciones:
 - Genera 4 criterios como máximo.
 - Crea 4 niveles de desempeño.
 - Usa descriptores observables, sin adjetivos vagos.
 - Evita mezclar dos variables en el mismo descriptor.
 - Añade ejemplos de errores frecuentes.
 - Devuelve:
 - 1) tabla para profesorado,
 - 2) versión resumida para alumnado,
 - 3) JSON con campos: criterio, nivel_1, nivel_2, nivel_3, nivel_4.

Plantilla JSON de rúbrica

- “ **Tarea:** exposición oral
Curso: 5.º Primaria / 2.º ESO / 1.º Bach.
Criterios: contenido, organización, lenguaje, interacción con audiencia
Niveles: inicial / básico / adecuado / avanzado
Regla: cada descriptor debe ser observable y no repetir el del nivel contiguo.

```
{
  "instrumento": "rubrica_analitica",
  "etapa": "ESO",
  "curso": "2",
  "materia": "Lengua Castellana y Literatura",
  "tarea": "texto argumentativo",
  "criterios": [
    {
      "id": "C1",
      "nombre": "Tesis y posicionamiento",
      "peso": 0.25,
```

```

    "niveles": {
      "1": "No formula tesis o es contradictoria.",
      "2": "Formula una tesis poco clara o incompleta.",
      "3": "Formula una tesis clara y pertinente.",
      "4": "Formula una tesis precisa, matizada y bien sostenida."
    }
  }
}
}

```

De la rúbrica a la lista o a la escala

Una ventaja real de la IA es la **transformación entre formatos**. Si una rúbrica ya está bien hecha, el modelo puede convertirla en checklist o escala sin rehacer el diseño desde cero. Por ejemplo:

- **De rúbrica a checklist:** extraer los mínimos exigibles.
- **De rúbrica a escala:** convertir niveles en anclas verbales o numéricas.
- **De observaciones cualitativas a categorías:** agrupar notas docentes por patrones.

Esta reconversión es útil cuando queremos mantener coherencia entre evaluación del profesorado, autoevaluación y coevaluación. Estudios recientes sobre feedback y rúbricas muestran que cuando la IA trabaja **anclada en una rúbrica dada por el docente**, la calidad del feedback mejora, pero siguen apareciendo inconsistencias de puntuación que exigen calibración humana.

De la generación de la rúbrica a su uso real en el aula

Uno de los errores más frecuentes al empezar a trabajar con IA en evaluación es pensar que el proceso termina cuando el modelo genera una rúbrica o una tabla. En realidad, ese es solo el punto de partida. El verdadero valor aparece cuando el docente transforma esa salida en un instrumento listo para usar dentro de su flujo habitual de evaluación.

La IA puede ayudar en todas las fases:

- diseño inicial,
- revisión de criterios,
- generación de formatos reutilizables,
- exportación a herramientas digitales,
- integración con LMS,



- creación de feedback automatizable,
- y preparación de versiones para alumnado y profesorado.

El proceso completo suele seguir esta secuencia:

1. Definir el objetivo de evaluación

El docente concreta:

- qué quiere evaluar,
- qué evidencia recogerá,
- qué criterios curriculares intervienen,
- y qué uso tendrá la rúbrica.

No es lo mismo una rúbrica para:

- autoevaluación,
- corrección rápida,
- evaluación compartida,
- o integración automática en una plataforma digital.

Ese objetivo condiciona el diseño posterior.

2. Redactar el prompt inicial

La calidad de la rúbrica depende en gran parte de la precisión del prompt. Cuanto más contexto reciba la IA, más útil será el borrador generado.

Ejemplo de prompt completo:

“ Actúa como especialista en evaluación competencial de Lengua Castellana en 2.º ESO.

Diseña una rúbrica analítica para evaluar un texto argumentativo.

Condiciones:

- máximo 4 criterios,
- 4 niveles de desempeño,
- descriptores observables,



- lenguaje comprensible para alumnado de 13-14 años,
- incluir errores frecuentes,
- alineada con este criterio curricular:
[pegar criterio].

Devuelve:

1. tabla para profesorado,
2. versión resumida para alumnado,
3. JSON reutilizable para LMS o hojas de cálculo.

3. Revisar pedagógicamente la propuesta de la IA

La IA puede generar rápidamente descriptores, pero el docente debe validar:

- si los niveles realmente son distinguibles,
- si los criterios miden lo importante,
- si el lenguaje es adecuado para la edad,
- y si los indicadores son observables.

Este paso es esencial porque muchos modelos tienden a:

- repetir ideas entre niveles,
- usar lenguaje excesivamente genérico,
- o mezclar varios indicadores en un mismo descriptor.

La IA acelera el diseño, pero la validación pedagógica sigue siendo humana.

4. Elegir el formato final de uso

Una misma rúbrica puede transformarse en distintos formatos según el contexto de aplicación.

Tabla para impresión o uso manual

La salida puede copiarse directamente a:

- un documento de texto,
- una plantilla PDF,
- o una ficha impresa para aula.

Es la opción más sencilla y útil para:

- observación directa,
- exposiciones orales,
- correcciones rápidas,
- o evaluación en papel.

Hoja de cálculo

La tabla generada por la IA puede pegarse en:

- Excel,
- Google Sheets,
- LibreOffice Calc.

Esto permite:

- calcular puntuaciones automáticamente,
- generar medias,
- visualizar progreso,
- filtrar criterios,
- crear gráficos,
- o reutilizar rúbricas entre grupos.

Ejemplo de estructura:

Alumno	C1	C2	C3	C4	Observaciones
ALU_01	3	2	4	3	Mejorar cohesión
ALU_02	2	2	3	2	Revisar tesis

A partir de aquí, la IA puede incluso analizar patrones globales del grupo.

Integración en LMS

Muchos LMS permiten importar rúbricas o criterios mediante tablas o estructuras JSON.

Por ejemplo:

- Moodle
- Google Classroom
- Canvas

- Additio

El JSON generado por IA puede servir como estructura base para:

- automatizar criterios,
- crear escalas,
- importar indicadores,
- o reutilizar rúbricas entre cursos.

Ejemplo:

```
{
  "nivel_1": "No existe estructura",
  "nivel_2": "La estructura es",
  "nivel_3": "La organización es",
  "nivel_4": "La estructura es clara y"
}
```

Este tipo de formato facilita además futuras automatizaciones.

5. Utilizar la rúbrica como base para corrección asistida con IA

Aquí es donde enlaza directamente con el apartado 4.4 que veremos más en detalle.

Una vez creada la rúbrica, el docente puede reutilizarla como marco estable para:

- analizar producciones,
- generar feedback,
- detectar patrones,
- o redactar observaciones coherentes.

La rúbrica deja de ser solo un instrumento de calificación y se convierte en una “estructura de diálogo” entre docente, alumnado e IA.

Por ejemplo:

1. El docente pega la rúbrica.
2. Añade un texto anonimizado.
3. La IA analiza únicamente según esos criterios.



4. Devuelve evidencias y sugerencias.
5. El docente valida el resultado.

Este enfoque reduce comentarios genéricos y mejora la coherencia del feedback.

6. Generar feedback reutilizable y personalizado

Una ventaja especialmente potente es reutilizar los resultados para crear distintos tipos de retroalimentación:

- feedback breve,
- versión motivadora,
- adaptación por edad,
- comentarios para familias,
- o recomendaciones de mejora.

Ejemplo:

“ Reescribe este feedback para alumnado de 5.º de Primaria con tono motivador y dos pasos concretos de mejora:
[pegar feedback]

La IA puede transformar automáticamente un comentario técnico del profesorado en una explicación comprensible para el alumnado.

7. Conservar la rúbrica como instrumento reutilizable

Cuando la rúbrica ya está validada, puede convertirse en:

- plantilla reutilizable,
- banco de instrumentos,
- estructura para futuras situaciones de aprendizaje,
- o sistema de evaluación común entre docentes.

Muchos equipos docentes están empezando a crear bibliotecas compartidas de rúbricas generadas y refinadas con IA, lo que reduce tiempo de diseño y mejora la coherencia entre materias y niveles.

El verdadero potencial de la IA en evaluación no está en que “corrija sola”, sino en que permite convertir instrumentos bien diseñados en sistemas de feedback más coherentes,



reutilizables, transparentes y pedagógicamente útiles.

3.3. Listas de cotejo

La lista de cotejo es un instrumento más simple que la rúbrica y sirve para comprobar si aparecen o no determinados elementos o si cumple o no una serie de requisitos. Es especialmente útil en tareas objetivas con requisitos concretos, productos estructurados, observación rápida de conductas o verificación de entregables. Su valor reside en la claridad y en la rapidez de uso, por lo que funciona muy bien en fases iniciales de aprendizaje, en actividades procedimentales o cuando queremos verificar la presencia de componentes mínimos.

Desde el punto de vista teórico, la lista de cotejo se sitúa en un nivel más descriptivo que valorativo. No explica tanto la calidad como la existencia de un requisito. Por ello, es muy útil cuando la prioridad es asegurar que el trabajo cumple unas condiciones básicas o elementos esenciales antes de pasar a una evaluación más rica.

Ejemplo práctico: lista de cotejo para un proyecto

- Incluye portada.
- Presenta índice.
- Cita al menos dos fuentes.
- El trabajo está redactado en lenguaje propio.
- Hay conclusiones personales.
- La presentación se entrega en plazo.

Ejemplo práctico: lista de cotejo para un trabajo escrito

- Tiene título.
- Incluye introducción.
- Desarrolla al menos tres ideas.
- Usa conectores textuales.
- Incluye conclusión.
- Cita al menos una fuente.

Ejemplo práctico: lista de cotejo para una presentación oral

- Mira al público.
- Habla en voz audible.
- Respeta el tiempo.
- No lee todo el texto.
- Explica las ideas principales.



La IA puede ayudar a convertir criterios demasiado ambiguos en ítems observables. Por ejemplo, una instrucción como “el alumno se expresa bien” puede transformarse en “usa vocabulario adecuado al tema”, “mantiene una secuencia lógica” y “pronuncia con claridad”. Eso mejora la fiabilidad del instrumento.

Este instrumento es especialmente útil en etapas iniciales, tareas cortas o actividades en las que se quiere evitar una carga de corrección excesiva. La IA puede ayudar a convertir un enunciado de tarea en una lista de verificación clara y breve.

Prompt para convertir una rúbrica en lista de cotejo

“ A partir de esta rúbrica, transforma solo los indicadores que sean verificables en una lista de cotejo sí/no.
Marca aparte los que NO conviene convertir a sí/no porque exigen gradación.
Devuelve la salida en dos bloques:

- Lista de cotejo aplicable
- Indicadores que deben quedarse en escala o rúbrica

“ A partir de esta rúbrica, genera una lista de cotejo de requisitos mínimos observables para uso rápido en aula. Mantén coherencia con los criterios y evita ítems redundantes.

3.4. Escalas de valoración

La escala de valoración es muy útil cuando queremos medir grado / calidad , frecuencia o intensidad de un comportamiento o desempeño. A diferencia de la lista de cotejo, no se limita a un sí o no, sino que introduce niveles graduales. Pueden ser rangos numéricos o verbales (“nunca-a veces-frecuentemente-siempre” o “insuficiente-aceptable-adecuado-excelente”), y es muy habitual en observación del trabajo en grupo, participación, hábitos de trabajo o evolución de determinadas competencias.

Son útiles cuando queremos observar procesos más abiertos que una simple comprobación sí/no. Teóricamente, la escala permite captar matices. No todos los aprendizajes se pueden reducir a presencia/ausencia, y hay procesos en los que interesa saber si algo ocurre rara vez, a veces o con frecuencia. La escala es, por tanto, un puente entre lo descriptivo y lo valorativo.

Ejemplo práctico: escala numérica para trabajo cooperativo

Indicador	1	2	3	4
Participa activamente				
Escucha a sus compañeros				
Aporta ideas útiles				
Respeto los turnos				

Ejemplo práctico: escala con rangos verbales para trabajo cooperativo

Indicador	Nunca	A veces	Frecuentemente	Siempre
Escucha a sus compañeros				
Aporta ideas útiles				
Respeto los turnos				
Cumple su parte del trabajo				

También puede construirse con niveles de logro:

- Insuficiente.



- Básico.
- Adecuado.
- Excelente.

Ejemplo de aplicación: en un proyecto de ciencias, la escala puede valorar si el alumnado participa activamente, si colabora, si gestiona bien el tiempo y si asume responsabilidades. La IA puede ayudar a redactar los indicadores para que sean observables, pero el docente debe decidir qué significa exactamente “frecuentemente” en ese contexto.

La IA puede ayudar a redactar indicadores observables, evitando términos vagos como “se implica mucho” o “trabaja bien”, que luego son difíciles de valorar con consistencia.

Prompt para crear una escala de valoración

“ A partir de este criterio y de estas evidencias, crea una escala de valoración de 1 a 4.
Ancla cada puntuación con frecuencia o autonomía observable.
No uses “bien”, “regular” o “mal”.
Usa fórmulas como:
1 = no lo realiza ni con ayuda
2 = lo realiza con mucha ayuda
3 = lo realiza con alguna ayuda
4 = lo realiza de forma autónoma y consistente

Prompt para convertir una rúbrica a escala de valoración

“ Convierte esta rúbrica en una escala de valoración de 4 niveles con anclas verbales: nunca / a veces / frecuentemente / siempre, o insuficiente / básico / adecuado / excelente según convenga a la tarea. Explica qué evidencia observable corresponde a cada nivel.

3.5. Registros de observación

El registro de observación es uno de los instrumentos más ricos para evaluar el proceso, no solo el producto. Permite anotar conductas, interacciones, dificultades, avances y decisiones pedagógicas, es decir, evidencias mientras el alumnado trabaja. Es especialmente valioso en situaciones prácticas, trabajo en taller, aprendizaje cooperativo, actividades orales o dinámicas de aula para valorar la interacción, participación, desempeño oral o trabajo práctico donde la evidencia no queda recogida en un documento final.

A diferencia de otros instrumentos, aquí importa tanto lo que se observa como el momento en que se observa. Teóricamente, este instrumento reconoce que aprender no siempre deja una huella escrita. Muchas veces, la evidencia aparece en la forma de participar, resolver problemas, explicar, justificar, corregir o colaborar. El registro de observación da valor a esos momentos efímeros que de otro modo se perderían.

Ejemplo práctico: registro breve de aula

- Fecha.
- Alumno o grupo.
- Tarea o actividad.
- Conducta o evidencia observada.
- Dificultad o incidencia detectada.
- Intervención docente.
- Decisión posterior.

Ejemplos de anotaciones reales:

“ Durante la resolución del caso, el grupo identifica correctamente el problema, pero no justifica la solución. Necesitan apoyo en argumentación.

“ Durante la resolución del caso, el grupo identifica bien el problema, pero no justifica la solución con argumentos científicos. Se propone andamiaje con preguntas guía en la siguiente sesión.



La IA puede ayudar a organizar registros extensos, agrupar observaciones por categorías o resumir patrones, pero la observación en sí debe ser humana y contextual. Si la IA interpreta por su cuenta comportamientos del alumnado, se corre el riesgo de sobregeneralizar o de convertir una anécdota en una conclusión injustificada.

Con IA, el registro puede transformarse después en patrones o categorías, pero la observación original debe ser del docente. La IA no debería sustituir el juicio profesional ni convertir pequeños rastros de aula en diagnósticos automáticos.

Prompt para registro de observación

“ Diseña un registro de observación para [Infantil/Primaria/ESO/FP] en la actividad [x].

Debe incluir:

- fecha
- contexto
- conducta o evidencia observable
- frecuencia o momento de aparición
- interpretación pedagógica breve
- siguiente decisión docente

Evita etiquetas clínicas o diagnósticas.

Devuelve una versión en tabla y otra para copiar a una hoja de cálculo.

“ Diseña un registro breve de observación para [actividad] con columnas de fecha, código del alumno/grupo, evidencia observada, dificultad detectada, intervención docente y decisión de seguimiento.