

Criterio	Descripción	1 (Deficiente)	2 (Aceptable)	3 (Bueno)	4 (Excelente)
1. Tiempo de respuesta	Velocidad en entregar la respuesta tras el envío del prompt.	> 3 min (muy lento)	1 min – 3 min (lento)	30 s – 60 s (rápido)	< 30 s (muy rápido)
2. Cumplimiento de formato	Ajuste a la estructura solicitada (lista enumerada, límite de palabras, idioma, etc.).	No respeta el formato ni estructura	Respeto parcialmente (p. ej., lista desordenada o excesiva longitud)	Casi cumple (lista clara, pero algún detalle fuera de formato)	Cumple al 100 % (lista numerada, longitud y estilo exactos)
3. Precisión y veracidad	Ausencia de “hallucinations”: sólo menciona recursos existentes y hace referencia a datos correctos.	Introduce varios recursos o datos ficticios	1–2 errores de recursos/datos	Sin recursos inventados, pero alguna pequeña incongruencia	Completamente libre de errores o invenciones
4. Relevancia y adecuación	Grado en que el itinerario se ajusta a la etapa, asignatura y duración solicitadas.	No cumple los requisitos principales (etapa, materia o duración)	Cumple uno de los tres requisitos	Cumple dos de los tres requisitos	Cumple plenamente etapa, materia y duración
5. Calidad del itinerario	Secuencia pedagógica, variedad y pertinencia de recursos, justificación de cada actividad.	Secuencia incoherente, sin justificación ni variedad	Pobre organización, justificaciones débiles	Buena coherencia y justificación aceptable	Excelente secuencia didáctica, justificaciones sólidas y recursos muy variados

Criterio	Descripción	1 (Deficiente)	2 (Aceptable)	3 (Bueno)	4 (Excelente)
6. Robustez de respuesta	Capacidad de siempre devolver un itinerario válido (sin “no sé” ni vacío).	A menudo falla o responde “No sé”	Ocasionalmente omite la propuesta	Raramente omite; siempre ofrece algo aunque imperfecto	Siempre genera itinerario válido y completo