# Diseño de instrumentos de evaluación curricular asistido por inteligencia artificial generativa

## Design of curricular evaluation instruments assisted by generative artificial intelligence

Fecha de recepción:
11 de diciembre 2025

Fecha de aprobación:
03 de febrero 2026

https://hdl.handle.net/20.500.14492/33210

Giselle Angélica Ñurinda-Montoya
Universidad Nacional, Costa Rica
giselle.nurinda.montoya@una.ac.cr
https://orcid.org/0009-0005-7910-4316

Kendall Antonio Ruiz Benavides
Universidad Nacional, Costa Rica
kendall.ruiz.benavides@una.ac.cr
https://orcid.org/0009-0004-7747-3700

Mariangelina Rodríguez Varela
Universidad Nacional, Costa Rica
mariangeli.rodriguez.varela@una.ac.cr
https://orcid.org/0009-0002-1589-736X

Gabriela María Gutiérrez López
Universidad Nacional, Costa Rica
gabriela.gutierrez.lopez@una.ac.cr
https://orcid.org/0009-0001-5979-359X

## Resumen

El uso de la inteligencia artificial generativa ejerce un impacto cada vez más significativo en la vida cotidiana, abarcando desde tareas simples hasta procesos altamente complejos. En este contexto, se consolida como una herramienta de apoyo valiosa para el diseño y la validación de instrumentos de evaluación educativa. El presente estudio tiene como objetivo diseñar instrumentos de evaluación curricular mediante la aplicación del modelo CIPP, la teoría de diseño de instrumentos fundamentada en la psicometría moderna de Messick y los principios de la ingeniería del *prompting*, utilizando la inteligencia artificial generativa como herramienta de apoyo metodológico. La investigación se basa en la metodología de Diseño y Desarrollo Basado en Teoría. Los resultados se presentan en forma de productos tangibles, entre los cuales se incluyen: (a) matriz de dimensiones, criterios e indicadores derivada del modelo CIPP; (b) definición del constructo de evaluación; y (c) el *framework* de *prompting* por componentes; este último producto está conformado por tres prompts. En conclusión, se muestra que la integración del modelo CIPP, la teoría psicométrica y la ingeniería del *prompting* permite diseñar prompts estructurados que guíen a la inteligencia artificial generativa a la generación de ítems válidos y coherentes, fortaleciendo la calidad y pertinencia del diseño de instrumentos de evaluación curricular.

**Palabras clave:** evaluación curricular, evaluación de programas, instrumentos de medición, inteligencia artificial, modelo CIPP.

## Abstract

The use of generative artificial intelligence is having an increasingly significant impact on everyday life, ranging from simple tasks to highly complex processes. In this context, it is establishing itself as a valuable support tool for the design and validation of educational assessment instruments. The present study aims to design curriculum assessment instruments by applying the CIPP model, Messick's modern psychometric-based instrument design theory, and the principles of prompting engineering, using generative artificial intelligence as a methodological support tool. The research is based on the Theory-Based Design and Development methodology. The results are presented in the form of tangible products, including: (a) a matrix of dimensions, criteria, and indicators derived from the CIPP model; (b) a definition of the assessment construct; and (c) the component-based prompting framework, which consists of three prompts. In conclusion, it is shown that the integration of the CIPP model, psychometric theory, and prompting engineering allows for the design of structured prompts that guide generative artificial intelligence to generate valid and coherent items, strengthening the quality and relevance of the design of curriculum assessment instruments.

**Keywords:** artificial intelligence, CIPP model, curriculum evaluation, measurement instruments, program evaluation.

## 1. Introducción

El diseño de instrumentos de medición educativa ha sido históricamente un proceso riguroso que exige tanto precisión conceptual como coherencia metodológica. Tradicionalmente, la generación de ítems y la validación de instrumentos dependen de la revisión teórica, la consulta a expertos y el análisis estadístico. No obstante, la irrupción de la inteligencia artificial generativa (IAG) y, en particular, de los modelos de lenguaje generativo, abre nuevas posibilidades para transformar estos procesos desde una perspectiva de inteligencia asistida.

En esta investigación, la IAG se entiende como un conjunto de sistemas capaces de producir contenido nuevo y original, a partir de datos y patrones previamente aprendidos. Dentro de esta categoría, los modelos de lenguaje de gran escala (*Large Language Models*, LLM) representan una de sus manifestaciones más relevantes, en este estudio se utilizan dichos modelos de lenguaje para el diseño de instrumentos. Estos modelos están diseñados para comprender y generar lenguaje natural, lo que permite un diálogo con la IAG a través de instrucciones escritas, conocidas como *prompts* (Korzynski et al., 2023; Schulhoff et al., 2024; Velásquez-Henao et al., 2023).

Desde una perspectiva educativa, la IAG se ha convertido en una herramienta emergente capaz de asistir al profesorado en tareas complejas de diseño y redacción mediante el uso de *prompts* o instrucciones textuales. En este contexto, la ingeniería de *prompting* surge como una disciplina clave que permite estructurar y formular adecuadamente los *prompts* para obtener resultados pertinentes, precisos y éticamente adecuados. Esta técnica no solo potencia la productividad docente, sino que también contribuye a mejorar la calidad de los instrumentos mediante la automatización guiada de tareas de análisis y creación de ítems, rúbricas o escalas de valoración.

Asimismo, la evaluación educativa constituye el marco esencial para comprender la función y el valor de los instrumentos diseñados con apoyo de IAG. En este sentido, el modelo CIPP (Contexto, Insumo, Proceso y Producto), propuesto por Stufflebeam y Shinkfield (2007), ofrece un enfoque integral que orienta la toma de decisiones en cada fase de un programa o proyecto educativo.

De forma complementaria, la teoría unificada de la validez de Messick (1995) plantea que esta no es una propiedad del instrumento, sino un juicio integrador sobre la adecuación, interpretabilidad y consecuencias de las inferencias que se derivan de sus resultados. Esta perspectiva demanda considerar evidencia de contenido, estructura interna, relación con otras variables y consecuencias sociales de la evaluación, garantizando que los instrumentos generados, incluso los creados con apoyo de la IAG, mantengan su rigor conceptual y ético.

Integrar ambos enfoques permite articular una visión amplia y actualizada de la evaluación educativa: mientras el modelo CIPP aporta una estructura para analizar el proceso evaluativo en sus distintas etapas, la propuesta de Messick refuerza la necesidad de asegurar la validez y la pertinencia de las inferencias. En conjunto, ambos modelos ofrecen un marco sólido para explorar cómo la inteligencia artificial generativa puede contribuir a la calidad, la equidad y la transparencia en el diseño y validación de instrumentos de evaluación educativa.

El objetivo de este estudio es diseñar instrumentos de evaluación curricular mediante la aplicación del modelo CIPP, la teoría de diseño de instrumentos fundamentada en la psicometría moderna de Messick y los principios de la ingeniería del *prompting*, utilizando la inteligencia artificial generativa como herramienta de apoyo metodológico.

## 2. Desarrollo

### 2. 1 Evaluación curricular como herramienta para la mejora educativa

La evaluación curricular constituye un proceso esencial dentro de los sistemas educativos contemporáneos, orientado a valorar de manera sistemática, continua y rigurosa la pertinencia, coherencia y eficacia de los planes de estudio. No se limita a un ejercicio técnico de medición, sino que representa una actividad reflexiva que busca garantizar la calidad, relevancia y equidad de los procesos educativos. En palabras de Carreño y Chadwick, citados por Contreras (2021), la evaluación curricular es «un proceso sistemático que valora el grado en que los medios, recursos y procedimientos permiten el logro de las metas en un sistema educativo; contempla la delimitación, obtención y elaboración de información útil para juzgar la posibilidad de tomar decisiones» (p. 9). Este concepto enfatiza la necesidad de utilizar la evaluación como un mecanismo para orientar decisiones de mejora, sustentadas en la evidencia y contextualizadas en las condiciones reales de las instituciones educativas.

La evaluación curricular, además de ser un proceso técnico, tiene una dimensión ética y formativa. Aránguiz (2020) la define como «un proceso sistemático y continuo que permite valorar la pertinencia del plan de estudio con el contexto, con sus necesidades, problemas y tendencias, así como los diferentes componentes de la realidad institucional» (p. 2). Este enfoque resalta que la evaluación debe responder



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

a las demandas del entorno, integrando una mirada holística sobre la relación entre la educación y la sociedad. Por tanto, una evaluación curricular efectiva no se reduce a medir resultados académicos, sino que considera el grado en que los currículos contribuyen al desarrollo humano, social y profesional de las personas aprendientes.

De igual manera, Inciarte y Canquiz (2001) sostienen que la evaluación curricular implica «un proceso de recolección, procesamiento e interpretación de información necesarios para conocer, comprender, emitir juicios y tomar decisiones sobre un currículo determinado que conduzca a su permanente mejoramiento y transformación» (p. 4). Bajo esta perspectiva, la evaluación es tanto un proceso científico como un acto reflexivo y político, pues busca comprender la realidad educativa y transformarla mediante acciones fundamentadas en el análisis de datos y la deliberación crítica. De ahí que el diseño de las estrategias e instrumentos de evaluación curricular deban incorporar criterios de validez, confiabilidad y pertinencia contextual, permitiendo la interpretación adecuada de los hallazgos.

Por otro lado, Cely y Quiñones (2022) destacan que la evaluación curricular posee un carácter investigativo, ya que está «orientada a determinar la eficiencia, coherencia, pertinencia y relevancia del currículo […] que permite construir canales de comunicación entre estudiantes, docentes, personal administrativo y el sector productivo respecto a las competencias de egreso que se desean obtener» (p. 153). Este planteamiento reafirma el valor de la evaluación como un proceso colaborativo que vincula a todos los actores educativos y promueve la construcción colectiva de conocimiento sobre la calidad y pertinencia del currículo. Así, evaluar el currículo supone no solo medir su eficacia, sino también generar espacios de diálogo y reflexión que contribuyan al desarrollo de una cultura institucional de mejora continua.

Stufflebeam y Shinkfield, citados por Mora (2004), establecen que toda evaluación debe cumplir con cuatro condiciones fundamentales: ser útil, factible, ética y exacta. Ser útil implica que los resultados contribuyan efectivamente a la toma de decisiones y a la resolución de problemas; ser factible demanda que los procedimientos sean viables y adecuados a los recursos disponibles; ser ética exige transparencia, respeto a los participantes y compromiso con la mejora; y ser exacta implica que los resultados sean válidos, confiables y objetivos. Estos principios deben guiar no solo la ejecución de la evaluación, sino también el diseño de los instrumentos que la hacen posible.

### 2. 2 El modelo CIPP: un enfoque integral para la evaluación curricular

El modelo de evaluación CIPP constituye una de las propuestas más influyentes y aplicadas en el ámbito de la evaluación educativa. Este modelo ofrece una estructura integral y holística que permite valorar de manera sistemática los distintos componentes de un programa o plan de estudio, orientando el proceso hacia la toma de decisiones fundamentadas y la mejora continua (Aziz et al., 2018; Pizano, 2014). Su enfoque trasciende la simple medición de resultados, integrando variables contextuales, recursos disponibles, procesos de implementación y productos obtenidos, lo cual proporciona una visión completa del funcionamiento curricular.

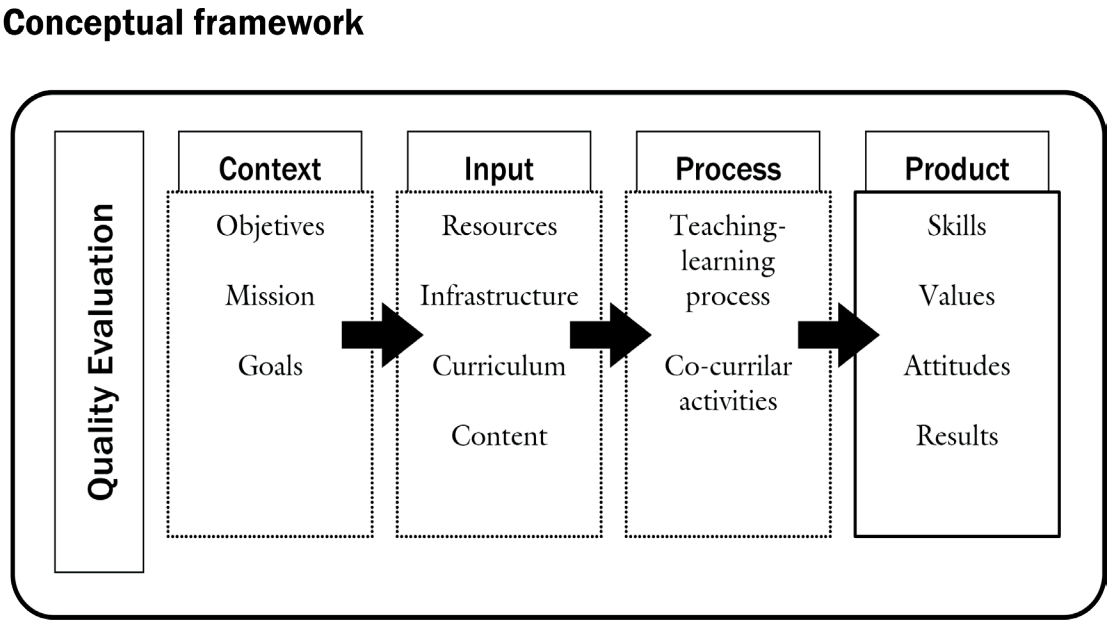
El modelo CIPP parte de la premisa de que la evaluación debe responder a preguntas esenciales: ¿qué se necesita? (Contexto), ¿qué recursos y estrategias se emplean? (Insumos), ¿cómo se desarrolla



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

el proceso educativo? (Proceso), y ¿qué resultados se logran? (Producto). De esta manera, permite analizar el currículo desde diferentes ángulos y etapas de desarrollo. En la dimensión de Contexto, la evaluación identifica las necesidades, problemas y oportunidades del entorno educativo. En la dimensión de Insumo, se examinan los recursos materiales, humanos y financieros. En la dimensión de Proceso, se analizan las actividades de implementación, la interacción entre personas docentes y estudiantes, la gestión de los recursos, así como las estrategias pedagógicas utilizadas. Finalmente, en la dimensión de Producto, se valoran los resultados del aprendizaje, las competencias adquiridas y el impacto del currículo en la formación integral del estudiantado (Aziz et al., 2018).

**Figura 1.** *Modelo de Evaluación Educativa CIPP*




*Nota.* Reproducido de Implementation of CIPP Model for Quality Evaluation at School Level: A Case Study (p. 195) por Aziz et al., 2018, Journal of Education and Educational Development.

Stufflebeam y Shinkfield, citados por Chamorro (2020), plantean que este modelo cumple cuatro funciones esenciales: (a) proporcionar datos útiles que den cuenta del mérito y validez del objeto evaluado, (b) posibilitar un examen exhaustivo en condiciones reales y dinámicas, (c) servir como guía para la toma de decisiones informadas, y (d) concebir la evaluación como un proceso continuo, sistemático y desarrollado por fases (p. 23).

Estas características hacen del modelo CIPP un marco adaptable a diversas realidades institucionales, ya que no impone un formato rígido, sino que ofrece criterios para orientar la evaluación en contextos complejos y cambiantes.

Desde esta perspectiva, el modelo CIPP se convierte en una herramienta indispensable para el diseño de instrumentos de evaluación curricular, pues cada dimensión ofrece un marco conceptual para la construcción de indicadores e ítems. Por ejemplo, en la dimensión de Contexto pueden diseñarse ítems



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

que exploren la alineación del currículo con las políticas nacionales o internacionales; en la dimensión de Insumo, indicadores relacionados con la formación docente o la disponibilidad de recursos; en la de Proceso, ítems que midan la pertinencia de las metodologías y estrategias didácticas; y en la de Producto, reactivos que valoren el logro de competencias y resultados de aprendizaje. Así, el modelo CIPP proporciona tanto una guía conceptual como metodológica para estructurar instrumentos válidos y coherentes con los propósitos de la evaluación.

### 2. 3 Diseño de instrumentos para la evaluación curricular

El diseño de instrumentos para la evaluación curricular constituye una fase crítica dentro del proceso de evaluación educativa, pues operacionaliza los constructos teóricos del modelo evaluativo en evidencias empíricas observables y medibles. Desde una perspectiva psicométrica contemporánea, la validez de un instrumento no se concibe como una propiedad intrínseca del instrumento, sino como un argumento unificado basado en la interpretación y el uso de los puntajes, siguiendo el enfoque propuesto por Messick (1989, 1995). Este autor amplió el concepto de validez integrando las dimensiones de contenido, criterio y constructo dentro de un marco de validez como inferencia argumentada, sustentada en evidencias empíricas y teóricas que justifican la interpretación de los resultados en contextos educativos específicos.

En este sentido, los *Standards for Educational and Psychological Testing* desarrollados por la *American Educational Research Association* (AERA), la *American Psychological Association* (APA) y el *National Council on Measurement in Education* (NCME) (2014) establecen que el diseño de instrumentos debe sustentarse en la coherencia entre el constructo teórico, el propósito de medición y las decisiones educativas derivadas de los resultados. Los estándares enfatizan cinco fuentes de evidencia para la validez: (a) contenido, (b) proceso de respuesta, (c) estructura interna, (d) relación con otras variables y (e) consecuencias del uso de la prueba, todas ellas aplicables a la evaluación curricular, donde se requiere demostrar tanto la pertinencia del contenido como la consistencia del modelo evaluativo.

El proceso de diseño de instrumentos de evaluación curricular se articula con los modelos de evaluación educativa como el CIPP, que permite valorar integralmente los componentes curriculares. En este marco, Karatas y Fer (2011) demostraron la aplicabilidad del modelo CIPP en el diseño de una escala para evaluar programas universitarios, validando una estructura de cuatro factores con alta confiabilidad ($\alpha = 0.91$). Esta estructura facilita vincular las dimensiones del contexto, insumo, proceso y producto con indicadores curriculares operativos, promoviendo una visión sistémica de la calidad formativa.

Desde una perspectiva latinoamericana, Carrillo-Cayllahua et al. (2023) desarrollaron un instrumento para la evaluación del currículo en una carrera universitaria, fundamentado en la relación entre el plan de estudio, el rendimiento académico y la permanencia estudiantil. Su enfoque metodológico evidenció la necesidad de instrumentos que integren tanto indicadores cuantitativos de desempeño como criterios cualitativos sobre la coherencia interna del currículo. Este tipo de aproximaciones responde al principio de triangulación planteado por Stufflebeam y Shinkfield (2007), que sugiere combinar evidencias de múltiples fuentes para fortalecer la validez del juicio evaluativo.



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

En el ámbito psicométrico, la construcción de ítems debe contemplar criterios de claridad, unicidad y pertinencia contextual, garantizando que cada ítem evalúe una sola acción o atributo (unicidad) (Ahmady et al., 2023). Dichos autores siguieron un proceso de seis etapas: búsqueda sistemática, estudio cualitativo, síntesis teórica, validación experta, análisis factorial exploratorio y fiabilidad, obteniendo un instrumento con cinco factores y una varianza explicada del 90 %, lo que ejemplifica el rigor requerido en la evaluación de constructos complejos como los procesos de aprendizaje o la gestión curricular en entornos virtuales. Asimismo, los resultados de Taureaux Díaz et al. (2016) en el diseño de un instrumento para la evaluación de la función de administración en el currículo de Medicina evidenciaron la importancia de la validación por jueces y la comprobación empírica de los resultados, alcanzando una correspondencia del 77 % entre las competencias declaradas y las observadas.

A nivel de validez de constructo y consistencia interna, la investigación de López et al. (2020) centrada en la docencia universitaria confirma que los índices psicométricos deben complementarse con análisis de correlación y estructura factorial confirmatoria, asegurando la estabilidad de los instrumentos en distintos contextos. Estos procedimientos coinciden con los lineamientos de los *Standards* (AERA, APA & NCME, 2014), que destacan la relevancia de la fiabilidad como condición necesaria, aunque no suficiente, para la validez.

En el campo de la evaluación asistida por inteligencia artificial (IA), los recientes aportes de Hernández-León y Rodríguez-Conde (2023) y Ramos-Armijos et al. (2023) muestran cómo los sistemas basados en IA pueden complementar la validación psicométrica tradicional mediante la analítica del aprendizaje y la modelización predictiva. Estas tecnologías permiten detectar sesgos, patrones de respuesta y coherencias entre dimensiones, fortaleciendo el análisis de la estructura interna y las inferencias derivadas. No obstante, tal como advierten Flores Contrera (2024) y Méndez-Mantuano et al. (2024), la integración de la IA en procesos evaluativos debe equilibrar el rigor técnico con los principios éticos de equidad, transparencia y protección de datos.

El diseño de instrumentos para la evaluación curricular debe conjugar la rigurosidad psicométrica clásica con una comprensión contextual e interpretativa del currículo. Esto implica construir instrumentos válidos, confiables y culturalmente pertinentes, que permitan obtener evidencias empíricas sólidas sobre la coherencia interna y externa de los programas académicos, contribuyendo a la toma de decisiones informadas para la mejora continua y la innovación educativa. La validez, en este contexto, se consolida como un proceso argumentativo continuo y multidimensional, sustentado en la convergencia entre teoría, evidencia y propósito evaluativo, tal como lo propuso Messick (1995) y lo reafirman los estándares contemporáneos de la evaluación educativa.

### 2. 4 Diseño y construcción de ítems

Un aspecto esencial en la construcción de instrumentos de evaluación es el diseño y construcción de ítems. Según Arias et al. (2014), la elección del formato de ítems depende del propósito del test y del tipo de habilidad o constructo que se desea medir, pudiendo requerir respuestas automatizadas, construidas, o una combinación de ambas.



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

Entre los formatos de respuesta más comunes tenemos los de elección múltiple o la selección de alternativas, también, formatos como ordenar, corregir, completar o construir respuestas completas, e incluso evaluar el desempeño en condiciones simuladas (Bennett, 1993, citado en Arias et al., 2014).

Una vez elegido el formato de respuesta se da paso al número de alternativas que se ofrecerán. Por ejemplo, en los test de rendimiento, suelen utilizarse formatos de verdadero/falso o de elección múltiple con entre tres y seis opciones, mientras que en los de personalidad predominan los formatos dicotómicos o las escalas graduadas. Entre estas últimas destacan la escala tipo Likert, el diferencial semántico y la escala análoga visual, seleccionadas según el nivel de precisión y el propósito de la medición. En todo caso, la elección del formato y del número de alternativas debe garantizar la validez y fiabilidad de los resultados obtenidos.

Al diseñar instrumentos de evaluación es importante considerar diversos aspectos relacionados con el formato de respuesta de los ítems. Según Arias et al. (2014), en primer lugar, los formatos dicotómicos, aunque reducen el tiempo de respuesta, tienden a generar distribuciones de puntuaciones sesgadas y menor varianza, por lo que requieren un mayor número de ítems para lograr precisión. En segundo lugar, los formatos graduados demandan decidir el número de puntos de la escala, que suele oscilar entre tres y nueve, cuidando que las etiquetas verbales sean lo suficientemente diferenciadoras. Además, utilizar un número impar de puntos permite incluir una opción neutral o intermedia, mientras que un número par obliga a una mayor discriminación en las respuestas.

Otro aspecto fundamental en el diseño y construcción de ítems es la redacción de estos. Deben ser claros, breves y comprensibles, evitando la ambigüedad y los inicios negativos y adaptándose al nivel lingüístico de la población objetivo. También es necesario evitar redundancias innecesarias, ya que aumentan el tiempo de lectura sin aportar valor, aunque cierta repetición controlada puede ser útil para verificar la consistencia de las respuestas. Finalmente, se recomienda considerar si es conveniente incluir ítems formulados tanto en sentido positivo como negativo para reducir sesgos de aquiescencia y afirmación, en todo caso, conviene evitar el uso con No al inicio de los ítems (Arias et al., 2014).

Una vez diseñados y elaborados los ítems junto con sus respectivas opciones de respuesta, se procede a la fase de revisión, etapa fundamental para prevenir errores que puedan comprometer la calidad del instrumento. En esta fase, se recomienda analizar los instrumentos desde distintas perspectivas, contando con la participación de expertos en el contenido, especialistas en psicometría y personas pertenecientes a la población objetivo. Los expertos en contenido deben garantizar que los ítems representen adecuadamente el constructo, los especialistas en medición, por su parte, revisan los aspectos metodológicos y la participación de personas de la población objetivo permite identificar posibles sesgos o lenguaje inapropiado, asegurando así la validez y equidad del instrumento.

### 2. 5 La ingeniería del prompting para el diseño de instrumentos

La ingeniería del *prompting* es un campo emergente que estudia la forma en que las personas diseñan y formulan instrucciones para comunicarse con sistemas de IAG (Velásquez-Henao et al., 2023; White et al., 2023; Schulhoff et al., 2024). En este estudio, las interacciones se realizan mediante Large Language Models (LLM), o grandes modelos de lenguaje, que son sistemas estadísticos entrenados con



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

volúmenes masivos de datos textuales capaces de leer, escribir y mantener diálogo en lenguaje natural (Maaz et al., 2025; White et al., 2023).

Más que una improvisación de indicaciones, esta disciplina propone un enfoque estructurado que une lenguaje, diseño y pensamiento estructurado, con el fin de lograr interacciones más precisas y productivas entre las personas y la inteligencia artificial generativa.

En este contexto, el *prompting* trasciende su función básica de dar instrucciones para constituirse como una metodología sistemática que orienta la construcción de conocimiento, la resolución de problemas complejos y el diseño de instrumentos mediante modelos de IAG. Desde esta perspectiva, la ingeniería del *prompting* se concibe como un fundamento metodológico, ya que enseña a formular *prompts* que no solo transmitan información, sino que también comuniquen intención, contexto y criterio de evaluación.

Dentro de estos fundamentos, uno de los marcos conceptuales más citados en la literatura es el modelo CLEAR, propuesto por Lo (2023), el cual organiza el diseño de *prompts* en cinco principios: *Concise* (preciso y breve), *Logical* (coherencia interna), *Explicit* (especificación clara del resultado esperado), *Adaptive* (capacidad de ajustar el *prompt* según el desempeño del modelo) y *Reflective* (evaluación continua para mejorar la instrucción). Este marco proporciona una estructura metodológica para optimizar la claridad, pertinencia y coherencia de las interacciones con modelos generativos (Lo, 2023).

Asimismo, la literatura describe enfoques iterativos como GPEI *(Goal–Prompt–Evaluation–Iteration)*, un ciclo metodológico que guía la formulación y refinamiento de prompts mediante cuatro etapas: *Goal*, donde se define la meta de la tarea; *Prompt*, donde se redacta la instrucción; *Evaluation*, donde se analiza críticamente la respuesta generada; e *Iteration*, donde se ajusta la instrucción para mejorar su precisión y utilidad. Este enfoque destaca que el *prompting* es un proceso de mejora continua orientado a la calidad metodológica de la interacción (Velásquez-Henao et al., 2023).

Según Schulhoff et al. (2024), los *prompts* son unidades de diseño que permiten controlar la coherencia y la relevancia de las respuestas generadas por los modelos de lenguaje. En el ámbito educativo, esta capacidad de control y orientación convierte al *prompting* en una estrategia que articula el conocimiento teórico con las capacidades generativas del modelo de IAG, posibilitando la co-construcción de ítems, indicadores o dimensiones de análisis (Velásquez-Henao et al., 2023).

Diversos autores coinciden en que el *prompting* no debe entenderse como una simple técnica espontánea, sino como una estrategia metodológica sustentada en principios comunicativos y cognitivos (White et al., 2023; Velásquez-Henao et al., 2023). La calidad de las respuestas de la IAG depende de la precisión semántica y de la claridad estructural del *prompt*. Por ejemplo, una instrucción vaga como «Haga preguntas para evaluar el programa de un curso» genera respuestas generales y poco alineadas con un marco evaluativo. En contraste, un prompt estructurado del tipo «Actúa como experta en evaluación curricular. Redacta ítems claros y unívocos basados en dimensiones y criterios previamente definidos, cuidando el estilo, la pertinencia contextual y la coherencia con el modelo evaluativo» produce resultados más pertinentes, consistentes y trazables.

Por ello, la ingeniería del *prompting* propone que cada interacción con la IAG sea diseñada de manera consciente, siguiendo criterios que garanticen validez, claridad y reproducibilidad.

En este sentido, el *prompting* puede entenderse como una mediación cognitiva y lingüística entre la persona y la inteligencia artificial. A través del lenguaje, las personas traducen su intención teórica en parámetros que la IAG puede interpretar y ejecutar.

De acuerdo con Lo (2023), el desarrollo de habilidades de *prompting* constituye una nueva forma de alfabetización académica, en tanto exige formular instrucciones precisas, coherentes y metacognitivamente justificadas para interactuar de manera eficaz con modelos de Inteligencia Artificial Generativa (IAG). Esta alfabetización no se limita al dominio técnico de la herramienta, sino que implica capacidades de reflexión crítica, control epistemológico, comunicación académica estructurada y evaluación de la validez de la información generada por los modelos.

Desde esta perspectiva, la práctica del *prompting* aporta rigor porque obliga a explicitar constructos, criterios y restricciones que tradicionalmente permanecían implícitos en los procesos de diseño de instrumentos; aporta trazabilidad porque deja un registro verificable de las decisiones metodológicas y de los ajustes sucesivos realizados en la interacción con la IAG y aporta transparencia porque hace visible el proceso de construcción, mostrando cómo cada componente del *prompt* se deriva de fundamentos teóricos y psicométricos previamente definidos. En conjunto, estos elementos refuerzan la importancia del *prompting* como herramienta científica para la construcción de instrumentos de evaluación educativa válidos, coherentes y reproducibles.

Con el tiempo, la práctica del *prompting* sigue volviéndose más común, generando la necesidad de sistematizar sus principios y componentes dentro de un modelo más amplio. Esa formalización da origen al *framework* de *prompting*, entendido como un modelo metodológico que organiza las partes y fases del proceso de interacción con la IAG.

De acuerdo con Schulhoff et al. (2024), un *framework* de *prompting* funciona como una estructura jerárquica que articula estrategias tales como la definición de roles, la contextualización de las tareas, la delimitación de las salidas esperadas y la iteración de resultados. Su propósito es garantizar precisión, coherencia y reproducibilidad en la generación de información por parte de la IAG.

En su dimensión estructural, el framework de *prompting* incluye componentes esenciales que orientan el diseño de las instrucciones. Entre los más importantes se encuentran: [Rol], que define la perspectiva o especialidad desde la cual la IAG debe responder; [Acción u objetivo], que especifica la tarea concreta a realizar; [Contexto], que ofrece los antecedentes necesarios para comprender la solicitud; [Audiencia], que determina el nivel de lenguaje y la profundidad de la respuesta; [Límites o restricciones], que establecen los criterios de control y formato; y [Estilo, Formato y Tono], que aseguran la coherencia comunicativa y el tipo de discurso que se espera obtener (White et al., 2023; Velásquez-Henao et al., 2023).

Cada uno de estos elementos cumple una función metodológica que permite controlar la calidad y la pertinencia del resultado, de modo que el *prompt* deja de ser una simple orden técnica para



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

convertirse en una decisión metodológica y comunicativa. Según Korzynski et al. (2023), la ingeniería del *prompting* implica una competencia estructurada que combina mediación cognitiva, precisión lingüística y control de la respuesta generada por la IA. En este sentido, la estructura del *prompt* puede entenderse como un mecanismo organizado de comunicación con la inteligencia artificial, en el que cada componente cumple una función específica para orientar y dar coherencia al resultado generado.

White et al. (2023) complementan esta visión al introducir los *prompt patterns*, o patrones de *prompting*, que funcionan como plantillas reutilizables para construir *prompts* en distintos contextos. Estos patrones —como el rol experto, el formato de salida o la iteración controlada— constituyen los bloques de construcción del *framework,* permitiendo estandarizar el proceso y asegurar su coherencia metodológica.

El *framework* de *prompting* puede considerarse un modelo de diseño cognitivo-lingüístico que integra tanto los aspectos técnicos como los metodológicos del *prompting*. Cada componente del *prompt* representa una decisión intencionada, similar a la formulación de ítems, criterios o indicadores en el diseño de instrumentos de evaluación (Velásquez-Henao et al., 2023).

Por esta razón, el *framework* de *prompting* constituye un modelo de mediación científica entre el lenguaje humano y la inteligencia artificial. Su aplicación en el diseño de instrumentos asistidos por IAG permite garantizar la validez conceptual, la trazabilidad del proceso y la consistencia teórica del conocimiento generado. De esta manera, la ingeniería del prompting y su *framework* metodológico se consolidan como nuevas herramientas educativas, al integrar el pensamiento humano con la capacidad generativa de la inteligencia artificial en un diálogo estructurado, ético y replicable.

## 3. Metodología

El presente estudio se fundamenta en la metodología de Diseño y Desarrollo Basado en Teoría (sus siglas en inglés TDD), un enfoque que busca generar conocimiento aplicable mediante la construcción de instrumentos teóricamente sustentados (Gregor y Hevner, 2013). En el contexto de la investigación educativa, este método implica el diseño y evaluación de soluciones innovadoras a problemas prácticos, guiadas por principios teóricos que explican y orientan el proceso (Reeves, 2006). En este caso, el producto consiste en la construcción de un *framework* de *prompting* para el diseño de instrumentos de evaluación curricular asistidos por IAG, apoyado en tres bases teóricas: la evaluación educativa, diseño de instrumentos y la ingeniería de *prompts*.

Desde la perspectiva de Gregor y Hevner (2013), la investigación de diseño no se limita a la construcción de un producto, sino que debe también contribuir a la teorización del proceso de diseño, articulando dos tipos de conocimiento: el conocimiento base (*knowledge base*, vinculado a las teorías y modelos existentes) y el conocimiento de diseño (*design knowledge*), que emerge de la práctica iterativa y reflexiva. En coherencia con esta visión, el presente estudio adopta un enfoque de desarrollo progresivo en tres fases secuenciales e interdependientes, descritas a continuación.



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

**Fase 1. Síntesis teórica y construcción del marco conceptual**

El objetivo de esta fase es integrar las tres áreas teóricas fundamentales que sustentan el modelo de diseño de instrumentos asistido por IAG:

a) La teoría de la evaluación educativa (modelo CIPP), que define el constructo, las dimensiones por evaluar y qué elementos se evalúan en cada una de ellas. (Stufflebeam y Shinkfield, como se cita en Chamorro, 2020; Aziz et al., 2018).

b) La teoría de diseño de instrumentos fundamentada en la psicometría moderna de Messick (1989, 1995) y los S*tandards for Educational and Psychological Testing* (AERA, APA y NCME, 2014), que establecen que todo instrumento debe reflejar de manera válida el constructo teórico que pretende medir. Esta teoría orienta la redacción de ítems, la selección de escalas de medición, la validación de contenido, el análisis factorial y comprobación de confiabilidad, asegurando que las evidencias empíricas respalden la interpretación de los resultados. Desde esta perspectiva, el diseño de instrumentos requiere evidencias empíricas y teóricas que garanticen la interpretación adecuada de los resultados y su aplicación contextualizada en procesos educativos.

c) La ingeniería del *prompting* en inteligencia artificial generativa, que define cómo comunicarse con un modelo de lenguaje para obtener respuestas consistentes y pertinentes (White et al., 2023; Velásquez-Henao et al., 2023).

En esta fase se desarrollan las siguientes actividades:

1. Revisión y sistematización del modelo CIPP como marco de referencia para la evaluación curricular.

2. Integración de los principios de validez unificada de Messick (1989) y los lineamientos de los *Standards* (AERA, APA y NCME, 2014) permite orientar la construcción de los instrumentos hacia cinco tipos de evidencias: de contenido, de proceso de respuesta, de estructura interna, de relaciones con otras variables y de consecuencias de uso. Estas guías constituyen el eje metodológico del diseño, complementando la lógica del modelo CIPP con fundamentos psicométricos verificables.

3. Análisis de marcos de ingeniería del *prompting* como *GPEI y CLEAR*, los cuales permiten la estructuración lógica de *prompts* educativos (White et al., 2023; Velásquez-Henao et al., 2023).

**Fase 2. Diseño del *framework* de *prompting* por componentes**

El objetivo de esta fase es diseñar un *framework* de *prompting* que, mediante una secuencia estructurada de *prompts*, sistematice el proceso de construcción de instrumentos de evaluación curricular asistido por IAG.



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

Las actividades de esta fase son:

1. Diseño de la estructura del modelo, estableciendo las etapas y principios que guían la construcción de instrumentos mediante IAG.

2. Definición de tres tipos de *prompts* para operacionalizar el proceso:

- ***Prompt* de definición conceptual y operacional:** tiene como finalidad establecer los fundamentos del instrumento mediante la delimitación de su objetivo general, el tipo de decisiones que sustentarán sus resultados, la población meta y sus características relevantes. Asimismo, operacionaliza el constructo a medir definiendo sus dimensiones, criterios e indicadores. La ejecución de este *prompt* requiere que se adjunten los insumos teóricos esenciales, tales como el marco teórico construido, documentos de especificación de dimensiones, criterios e indicadores y demás referentes conceptuales que aseguren la validez de contenido y constructo.

- ***Prompt* de construcción de ítems y formato del instrumento:** este *prompt* tiene como objetivo operacionalizar las dimensiones, criterios e indicadores predefinidos en la construcción concreta de los ítems y la estructura formal del instrumento. Su ejecución debe garantizar la generación de un cuestionario coherente, válido y listo para su aplicación, siguiendo estos criterios:

    - Cantidad de ítems:

        - El instrumento debe incluir un mínimo de 10 ítems en total.

        - Cada dimensión debe estar representada por, al menos, tres ítems. La cantidad final de ítems dependerá del número de dimensiones por evaluar. Si existe una dimensión que debe de ser más representada (puede depender de la población a la que va dirigida u otros factores) debe incluir más ítems. Es importante considerar que, según el objetivo de la evaluación, una dimensión puede estar representada por un mayor número de ítems que otra.

    - Redacción de ítems: la construcción debe regirse por los criterios de claridad (comprensión unívoca), unicidad (un solo concepto por ítem) y pertinencia contextual (relevancia para la población y el entorno evaluado). La redacción debe ser breve, sencilla, neutra, y evitar redundancias, negaciones, juicios de valor y sesgos. El lenguaje y la estructura sintáctica deben ser accesibles para la población meta.

    - Formato de respuesta: se debe especificar el formato de respuesta para los ítems (por ejemplo, escala Likert, elección múltiple, dicotómicas, diferencial semántico), el cual debe ser coherente con el objeto de estudio y el tipo de datos requeridos.

    - Estructura final del instrumento: la salida debe incluir:

        - Un título representativo del instrumento.



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

- Instrucciones generales que contengan: el objetivo de la evaluación, las condiciones de aplicación, características contextuales (institucionales, organizacionales o situacionales) y un espacio para el consentimiento informado.

- ***Prompt* de revisión y validación de instrumento:** solicita a la IAG la evaluación del instrumento generado en cuanto a claridad, sesgo, coherencia y alineación con los objetivos teóricos. Se debe evaluar con los siguientes criterios: El ítem tiene una redacción clara. El ítem utiliza un vocabulario fácil de comprender para la población objetivo. El ítem evalúa una sola acción (unicidad). El ítem corresponde a la dimensión que se pretende medir. El ítem aporta información útil para valorar el constructo. El contenido del ítem es pertinente al contexto de aplicación del instrumento. El ítem mantiene congruencia con la dimensión. Las opciones de respuesta son coherentes con la redacción del ítem. Existe correspondencia entre el ítem y la dimensión que se mide. El ítem es preciso sin ser restrictivo. El ítem está redactado de forma precisa respecto al objetivo del instrumento.

**Fase 3. Aplicación del *framework* de *prompting***

El objetivo de esta fase es comprobar la aplicabilidad del *prompt* mediante una prueba que demuestre su funcionalidad en un caso real de diseño de instrumento educativo.

Las actividades que pertenecen a esta fase son:

1. Selección de un caso práctico: diseño de un cuestionario para evaluar la percepción de docentes sobre la integración de la IAG en procesos curriculares.

2. Aplicación del *prompt*, documentando la interacción, las respuestas generadas por la IAG y las modificaciones realizadas.

3. Validación de los ítems producidos por la IAG por personas expertas en currículum y evaluación educativa.

## 4. Resultados y discusión

Los resultados del estudio se presentan como los productos tangibles generados mediante la aplicación de la teoría y la construcción de un *framework* de *prompting* por componentes, demostrando la viabilidad del modelo propuesto. Los resultados se organizan siguiendo el flujo de la investigación.

Estos resultados se muestran de manera integrada, dado que las distintas etapas del proceso metodológico convergieron en la obtención de productos complementarios que configuran un marco unificado de diseño asistido por IAG. Dichos productos son:

a) Matriz de dimensiones, criterios e indicadores derivada del modelo CIPP y definición del constructo de evaluación.



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

b) El *framework* de *prompting* por componentes, que operacionaliza las teorías de diseño de instrumentos y de *prompting* para la generación de ítems válidos, confiables y coherentes con el propósito evaluativo.

El primero corresponde al constructo por evaluar y la matriz de operacionalización del modelo CIPP, que traduce sus dimensiones teóricas en criterios e indicadores observables aplicables, a partir del caso práctico para ejemplificar la evaluación curricular, se seleccion***ó*** el Programa de Estudio de Biología del IV Ciclo del Ministerio de Educación Pública (MEP) de Costa Rica.

Esta estructura conceptual permitió organizar la información de forma jerárquica y definir los elementos clave del constructo por evaluar.

Constructo evaluativo identificado:

«El propósito del instrumento es evaluar la calidad integral del Programa de Biología a partir del análisis de su contexto, insumos, procesos y productos, de manera sistemática, contextualizada y participativa. Se busca valorar la pertinencia, coherencia y efectividad del programa, considerando tanto los factores institucionales y pedagógicos como los resultados obtenidos, con el fin de identificar fortalezas, áreas de mejora y el grado de contribución al desarrollo de competencias científicas, ambientales y socioeducativas del estudiantado»**.**

Con base en este constructo, la matriz de la tabla 1 establece las siguientes relaciones entre dimensiones, criterios e indicadores.

**Tabla 1.** *Matriz de operacionalización del modelo CIPP aplicado al programa de Biología*

| Dimensión | Criterio | Indicador |
|---|---|---|
| Contexto | Relevancia social y científica del programa | Nivel en que el programa de Biología aborda problemas científicos y ambientales actuales. |
| Contexto | Alineación con políticas educativas | Coherencia con el currículo nacional e internacional. |
| Contexto | Pertinencia para el desarrollo estudiantil | Relación del currículo con habilidades para educación superior y el mundo laboral. |
| Contexto | Consideración de diversidad y equidad | Relación del currículo con habilidades para educación superior y el mundo laboral. |
| Insumo | Adecuación de los recursos didácticos | Disponibilidad y calidad de materiales educativos. |
| Insumo | Formación y actualización docente | Nivel de capacitación de los docentes. |
| Insumo | Infraestructura y equipamiento | Nivel de satisfacción con el espacio áulico que cumple con las características para la ejecución del programa de Biología. |
| Proceso | Coherencia curricular | Coherencia entre ejes temáticos, criterios de evaluación y situaciones de aprendizaje. |
| Proceso | Secuencia curricular | Profundidad y secuencia de los ejes temáticos relacionados con los conocimientos y habilidades descritas. |



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

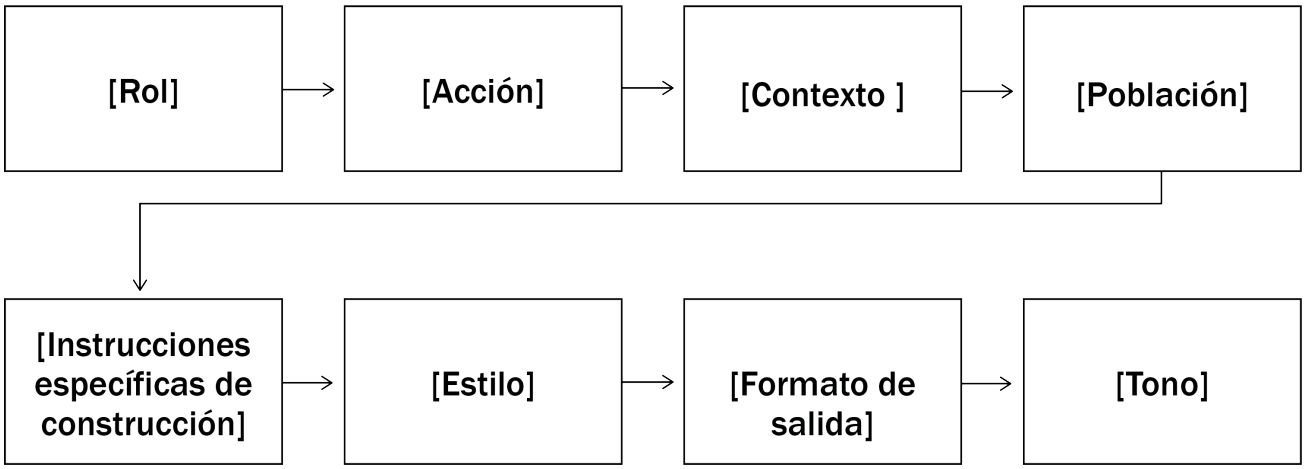
| | | |
|---|---|---|
| Proceso | Secuencia curricular | Percepción sobre la metodología de indagación sugerida para la mediación pedagógica (focalización, exploración, contrastación y aplicación). |
| Producto | Desempeño académico | Resultado de las evaluaciones académicas. |
| Producto | Impacto de la continuidad académica | Interés del estudiantado en carreras científicas. |
| Producto | Repitencia y convocatorias | Incidencia de los resultados de ampliación en la permanencia estudiantil. |

*Nota.* Esta matriz constituyó la base semántica del diseño instrumental y fue el insumo directo para el desarrollo de los prompts en la siguiente etapa.

El segundo consistió en la integración de los principios de la Teoría de Validez de Messick (1989) y los *Standards for Educational and Psychological Testing* (AERA, APA y NCME, 2014) dentro de la estructura de *prompting*.

A continuación, se presenta un modelo *framework* de *prompting* con tres componentes fundamentales, cada uno con una función específica dentro del flujo de diseño de instrumentos con IAG. Estos *prompt* siguen una estructura, la cual está representada en la Figura 2. Esta propuesta organiza sistemáticamente la interacción con el modelo de lenguaje para guiar la construcción metódica del instrumento.

**Figura 2.** *Estructura de los prompts*




*Nota.* Esta propuesta organiza sistemáticamente la interacción con el modelo de lenguaje para guiar la construcción del instrumento.

### Componente 1: plantilla para el *prompt* de definición conceptual y operacional

Este *prompt* tiene como propósito sentar las bases teóricas y estructurales del instrumento. Su función es guiar a la IAG en el análisis de los documentos teóricos esenciales (marco teórico, especificaciones de dimensiones, criterios e indicadores, entre otros) para definir con precisión el constructo a medir. El resultado es la operacionalización del constructo en dimensiones, criterios e indicadores medibles, garantizando la validez de contenido y constructo desde la fase inicial. A continuación, se especifica la redacción requerida para cada componente del *prompt*:



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

- [Rol]: aquí se debe redactar siempre el área de expertise específica y el modelo o marco teórico que guiará el análisis. Establece la «personalidad» experta de la IA.

- [Acción]: aquí se redacta la tarea central que debe realizar la IA. Se le debe instruir para que lea, analice y sintetice los documentos teóricos necesarios para la construcción del instrumento. La acción debe ser clara y dirigida hacia el objetivo de operacionalizar el constructo.

- [Contexto]: aquí se debe proporcionar la justificación y el propósito general del instrumento. Se debe definir el constructo a medir, el tipo de decisiones que se tomarán con los resultados, la población meta y sus características relevantes. Este elemento da sentido y dirección a toda la tarea.

- [Población]: aquí se define para quién está destinado el instrumento, así como las características específicas (nivel educativo, edad, género, país de procedencia) y, por ende, el lenguaje y el nivel de profundidad del análisis. Esto ajusta el tono y el enfoque de la salida de la IA (por ejemplo, docentes, estudiantes, para expertos técnicos o para tomadores de decisiones no especializados).

- [Instrucciones específicas de construcción]: se detallan los «cómo» de la tarea. Se deben listar criterios metodológicos clave, como priorizar evidencias observables y medibles, evitar solapamientos entre dimensiones, o garantizar que los indicadores sean alcanzables. Son las reglas de construcción.

- [Estilo]: aquí se especifica el tipo de lenguaje y terminología que debe emplear la IA en su respuesta (por ejemplo, técnico, académico, divulgativo).

- [Formato de salida]: aquí se define de manera explícita y estructurada cómo debe presentar la información la IA. Se debe solicitar un formato claro que facilite la revisión, como un informe con una tabla resumen que relacione las dimensiones, criterios e indicadores.

- [Tono]: aquí se establece la actitud o el matiz emocional de la comunicación (por ejemplo, formal, crítico-constructivo, neutral, persuasivo)

La Figura 3 presenta un ejemplo aplicado de la operacionalización del *prompt* de definición conceptual, materializando su estructura teórica en una instrucción aplicable. Esta ejemplificación visual comprobada, permite comprender la transición entre el marco abstracto y la construcción de un instrumento de evaluación válido y alineado con los objetivos de diagnóstico curricular para el programa de Biología.



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

**Figura 3.** *Prompt de definición conceptual*

**[Rol]**: actúa como un experto en evaluación curricular basado en el modelo CIPP.
**[Acción]**: lee los documentos adjuntos correspondientes al programa de Biología y el documento de las dimensiones, criterios e indicadores. Analiza las dimensiones, criterios e indicadores que servirán como base para un instrumento de evaluación del programa de Biología.
**[Contexto]**: el propósito del instrumento es evaluar la calidad integral del Programa de Biología a partir del análisis de su contexto, insumos, procesos y productos de manera sistemática, contextualizada y participativa la pertinencia, coherencia y efectividadsinbold del programa, considerando tanto los factores institucionales y pedagógicos como los resultados obtenidos. Con el fin de identificar fortalezas, áreas de mejora y su grado de contribución al desarrollo de competencias científicas, ambientales y socioeducativas del estudiantado.
**[Población]**: dirigido a docentes de instituciones públicas de educación media en Costa Rica.
**[Instrucciones específicas de construcción]**: prioriza la identificación de evidencias observables y medibles en los documentos para operacionalizar los indicadores.
**[Estilo]**: redacción técnica y precisa, siguiendo terminología educativa.
**[Formato de salida]**: presenta un informe del análisis que has hecho de las dimensiones, criterios e indicadores. Incluye en el informe una tabla resumen que relacione cada dimensión CIPP con sus criterios e indicadores derivados del análisis documental.
**[Tono]**: formal y descriptivo.

*Nota.* Elaboración propia.

### Componente 2: Plantilla para el *prompt* de construcción de ítems y formato del instrumento

Este *prompt* tiene como propósito fundamental guiar el diseño de ítems y del instrumento. A partir del marco de la validez de Messick, se operacionaliza la teoría mediante instrucciones específicas destinadas a generar evidencia de validez de contenido. Para ello, se indica a la IAG que los ítems deben derivarse directamente y cubrir de manera exhaustiva la matriz de operacionalización del modelo CIPP.

Asimismo, se especifican directrices para la estructura interna del instrumento, con el fin de que la IAG garantice una distribución equilibrada de ítems entre las dimensiones CIPP y mantenga coherencia en las escalas de respuesta. Además, se establecen criterios de redacción que priorizan la claridad y la unicidad de cada ítem. Las siguientes son las especificaciones de la redacción para cada elemento del *prompt*:

- [Rol]: aquí se debe redactar el perfil de un experto en psicometría y metodología de la investigación, con dominio en el diseño de instrumentos y en el marco teórico específico del constructo.

- [Acción]: aquí se instruye a la IA para que genere los ítems y la estructura del instrumento basándose directamente en las dimensiones, criterios e indicadores predefinidos. La acción debe ser directa: «Redacta», «Construye», «Diseña el instrumento».



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

- [Contexto]: aquí se reitera el propósito final del instrumento y se enfatiza la necesidad de alinear cada ítem con el constructo y las dimensiones teóricas. Se define qué se va a medir (percepciones, conocimientos, comportamientos) y en qué contexto.

- [Población]: aquí se especifica la población que responderá el instrumento. Esto es crucial para adecuar el lenguaje, la complejidad sintáctica y la contextualización de cada ítem.

- [Instrucciones específicas de construcción]: aquí se detallan los criterios psicométricos fundamentales para la redacción de ítems. Se deben incluir reglas como la claridad, unicidad, pertinencia contextual, neutralidad, y la prohibición de negaciones, juicios de valor o sesgos. También se especifican parámetros cuantitativos como el número mínimo de ítems por dimensión.

- [Estilo]: aquí se define el estilo lingüístico de los ítems y las instrucciones. Se debe especificar el tiempo verbal (por ejemplo, presente), el tipo de lenguaje (inclusivo, accesible) y el registro (formal, técnico).

- [Formato de salida]: aquí se describe de manera minuciosa la estructura final que debe tener el instrumento generado. Esto incluye los elementos formales (título, instrucciones, consentimiento informado) y la disposición específica de los ítems (agrupados por dimensión, en una tabla, con las opciones de respuesta claramente definidas y sus valores numéricos).

- [Tono]: aquí se establece la actitud general del instrumento hacia la población meta, que debe ser neutral, respetuosa y profesional, para fomentar la sinceridad y minimizar la deseabilidad social.

La Figura 4 ejemplifica la aplicación del *prompt* de construcción de ítems y formato del instrumento, donde la operacionalización teórica se materializa en un cuestionario estructurado y listo para su aplicación. En esta figura se visualiza cómo las dimensiones e indicadores predefinidos se traducen en ítems concretos, redactados bajo estrictos criterios psicométricos de claridad, unicidad y pertinencia y se organizan en una tabla con su escala de respuesta correspondiente. Asimismo, se muestra la incorporación de todos los elementos formales de un instrumento válido y confiable: título, instrucciones contextualizadas, consentimiento informado y una estructura lógica que guía a la población meta del instrumento.



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

**Figura 4.** *Prompt de construcción de ítems y formato del instrumento*

**[Rol]**: actúa como una doctora en Evaluación Educativa, experta en psicometría y especialista en el modelo CIPP.
**[Acción]**: redacta ítems coherentes a partir de las dimensiones, criterios e indicadores definidos en el documento.
**[Contexto]**: el instrumento busca evaluar percepciones sobre la implementación del programa de Biología, manteniendo la pertinencia y coherencia teórica con las dimensiones CIPP.
**[Población]**: dirigido a docentes del Ministerio de Educación Pública en educación media en Costa Rica.
**[Instrucciones específicas de construcción]**: la construcción de los ítems debe tener los criterios claros (comprensión unívoca), unicidad (un solo concepto por ítem) y pertinencia contextual (relevancia para la población y el entorno evaluado). La redacción de los ítems debe ser corta, sencilla y neutra, el lenguaje debe ser acorde a la población meta cuidando el nivel de dificultad de la lectura, evitar la redundancia, evitar el inicio con la palabra No, evitar que esté en negativo, evitar juicios de valorar, cuidar la estructura sintáctica y que sea acorde a las opciones de respuesta.
**[Estilo]**: redacción con verbos en tiempo presente y con un lenguaje inclusivo, adecuado para la población meta.
**[Formato de salida]**: instrumento que contenga título, las indicaciones del instrumento que deben de contener el objetivo, condiciones, características y factores propios del entorno institucional, organizacional o situacional desde los cuales se realiza la evaluación y consentimiento informado. Posteriormente que aparezca cada dimensión sus criterio e indicadores, agrega una tabla con dos columnas que: contenga 5 ítems por cada dimensión y un listado de las opciones de respuesta para cada ítem, las opciones de respuesta organízalas dentro de la tabla en una lista con viñetas de mayor a menor con la descripción y entre paréntesis el valor numérico de una escala de calidad con 5 categorías para cada ítem.
**[Tono]**: formal, neutral y respetuoso.

*Nota.* La Figura 4 muestra cómo el *prompt* transforma la operacionalización teórica en un cuestionario estructurado, con ítems claros y pertinentes, organizados con su escala de respuesta e incorporando todos los elementos formales de un instrumento válido.

### Componente 3: Plantilla para el *prompt* de revisión y validación de instrumento

Este *prompt* tiene como objetivo realizar una evaluación psicométrica rigurosa del instrumento generado, actuando como un primer filtro de validación de contenido. Su función es identificar problemas de redacción, sesgos, falta de coherencia o desalineaciones con el constructo teórico en cada ítem. El resultado es un informe detallado que justifica el veredicto sobre cada criterio y proporciona recomendaciones accionables para refinar el instrumento antes de su aplicación piloto o juicio de expertos humanos. A continuación, se detallan las especificaciones de redacción para cada elemento del *prompt*:



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

- [Rol]: aquí se debe redactar el perfil de un especialista en psicometría y métodos de evaluación, con dominio en los marcos teóricos que sustentan la validez (Teoría Clásica de los Test (TCT), Teoría de Respuesta al Ítem (TRI), entre otros). Este rol garantiza que la evaluación se realice con rigor metodológico.

- [Acción]: aquí se instruye a la IA para que realice una evaluación criterial, ítem por ítem. La acción debe ser analítica y deliberativa: «Evalúa», «Emite un veredicto», «Justifica tu decisión». Se debe enfatizar en la dicotomía Sí/No y en la obligatoriedad de proporcionar una justificación técnica y sugerencias de mejora.

- [Contexto]: aquí se debe especificar el instrumento bajo revisión, su propósito y la población a la que va dirigido. Es crucial mencionar que la evaluación se enmarca en un proceso de validación formal (por ejemplo, validez de contenido basada en TCT/TRI) para contextualizar la severidad del análisis.

- [Población]: aquí se define quiénes utilizarán el reporte de la IA (investigadores, comité de validación). Esto ajusta el nivel de tecnicismo del lenguaje en las observaciones, que debe ser preciso, pero comprensible para el destinatario.

- [Instrucciones específicas de construcción]: aquí se detalla la metodología de evaluación. Se deben listar explícitamente los criterios psicométricos a evaluar (claridad, unicidad, pertinencia, coherencia, entre otros). Es fundamental instruir a la IAG para que evalúe la redacción original y que sus sugerencias sean concretas y estén acotadas a una columna específica, sin reescribir los ítems en esta fase.

- [Estilo]: aquí se exige un estilo de redacción técnico y preciso, donde cada observación debe estar directamente anclada a uno de los criterios predefinidos, evitando generalidades o impresiones subjetivas.

- [Formato de salida]: aquí se debe exigir un formato que permita una revisión sistemática y eficiente. Una tabla es ideal, con columnas que desglosen el ítem, el criterio evaluado, el veredicto, la justificación técnica y la recomendación. Este formato facilita la posterior corrección del instrumento.

- [Tono]: aquí se establece que la actitud debe ser imparcial, fundamentada y orientada a la mejora. El tono debe ser «crítico-constructivo», lo que significa señalar deficiencias de forma objetiva y siempre proponer una vía de solución, fomentando el mejoramiento continuo del instrumento.

La Figura 5 ilustra la implementación del *prompt* de revisión y validación, donde el instrumento es sometido a un escrutinio psicométrico sistemático. En esta figura se ejemplifica cómo cada ítem es evaluado de forma independiente contra una lista exhaustiva de criterios de calidad, como la claridad, la unicidad y la coherencia con la dimensión asignada. La salida se estructura en una tabla de análisis que no solo emite un veredicto (Sí/No) para cada criterio, sino que también aporta una justificación



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

técnica y una recomendación concreta para los ítems que no cumplen, demostrando así un proceso de evaluación crítico-constructivo orientado a la optimización final del cuestionario.

**Figura 5.** *Prompt de revisión y validación*

**[Rol]**: actúa como un experto en evaluación psicométrica y análisis de validez de contenido, con amplia experiencia en la aplicación de la Teoría Clásica de los Test (TCT) y la Teoría de Respuesta al Ítem (TRI).
**[Acción]**: evalúa cada ítem del instrumento contra una lista de criterios predefinidos. Para cada criterio, debes emitir un veredicto dicotómico (Sí/No) y proporcionar una observación crítica-constructiva que justifique tu decisión y, en caso negativo, ofrezca una sugerencia de mejora concreta. Recuerda el principio de unicidad: cada ítem debe evaluar una sola acción o concepto.
**[Contexto]**: estás participando en la revisión y validación de contenido de un "Instrumento de evaluación curricular" para el programa de Biología a nivel de secundaria dirigido a docentes del Ministerio de Educación Pública. La validación se enmarca en los modelos de la TCT y la TRI.
**[Población]:** el resultado será revisado por investigadores y expertos en educación, por lo que el lenguaje debe ser técnico, pero claro.
**[Instrucciones específicas de construcción]**: evalúa estrictamente la redacción original de los ítems, sin proponer modificaciones en esta fase, solo sugerencias en la columna correspondiente. Aplica los siguientes criterios de forma independiente para cada ítem: 1. El ítem tiene una redacción clara. 2.El ítem utiliza un vocabulario fácil de comprender para la población objetivo.3. El ítem evalúa una sola acción (unicidad).4. El ítem corresponde a la dimensión que se pretende medir.5. El ítem aporta información útil para valorar el constructo.6. El contenido del ítem es pertinente al contexto de aplicación del instrumento. 7. El ítem mantiene congruencia con la dimensión. 8.Las opciones de respuesta son coherentes con la redacción del ítem.9. El ítem es preciso sin ser restrictivo.10. El ítem está redactado de forma precisa respecto al objetivo del instrumento.
**[Estilo]**: redacción técnica y precisa, con observaciones específicas y basadas en los criterios para cada ítem.
**[Formato de salida]**: presenta los resultados en una tabla por cada dimensión que incluya el análisis exhaustivo de todos los ítems. La tabla debe contener las siguientes columnas: Ítem: (texto del ítem evaluado) - Criterio(texto de los criterios por evaluar para cada ítem) - Cumple (Sí/No) (Veredicto dicotómico) -Observación (Justificación técnica de tu veredicto) - Recomendación (Sugerencia de mejora concreta en caso de ser "No").
**[Tono]**: Analítico, profesional y crítico-constructivo. El foco está en la mejora del instrumento.

*Nota.* La Figura 5 muestra cómo el *prompt* somete el instrumento a una revisión psicométrica sistemática, evaluando cada ítem según criterios de calidad y ofreciendo un análisis con justificación técnica y recomendaciones de mejora.



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

Estos *prompts* se obtuvieron tras un proceso de refinamiento especialmente en el elemento llamado «formato de salida», realizando algunas pruebas iterativas y sucesivas con la IAG. Este procedimiento permitió ajustar progresivamente las instrucciones hasta alcanzar respuestas deseadas, consistentes y alineadas con los propósitos evaluativos del instrumento. El conjunto de *prompts* optimizados demostraron la capacidad para guiar a la IAG en la generación de un instrumento con **ítems válidos** tal como se observa en la figura 6.

**Figura 6.** *Resultado del LLM del prompt revisión y validación*

✅ **Resumen General**

- Total de ítems evaluados:20
- Ítems que cumplen todos los criterios: 17
- Ítems con observaciones por unicidad o precisión: 3

✅ Conclusión técnica

El instrumento presenta alta validez de contenido (ICV ≥ 0.90 esperado) y buena coherencia interna teórica según el modelo CIPP. Sin embargo, se recomienda ajustar o dividir los ítems que integran más de un constructo, para garantizar el principio de unicidad y mejorar la discriminación ítem-factor de futuras fases de calibración (TCT o TRI)

*Nota.* Instrumentos con ítems válidos.

El resultado de este estudio fue la construcción de un instrumento completo asistido por IAG, compuesto por ítems tipo Likert organizados según las cuatro dimensiones del modelo CIPP. El proceso demostró que el *framework* de *prompting* propuesto fue adecuadamente diseñado, sustentado en teoría educativa y psicométrica, permitiendo a la IAG generar productos coherentes, reproducibles y evaluables.

Se confirma la viabilidad empírica del uso de la IAG como asistente metodológico en el diseño de instrumentos de evaluación curricular. El modelo de *framework* propuesto funcionó como un mediador epistemológico entre la teoría educativa (modelo CIPP) y la práctica psicométrica, garantizando coherencia conceptual y eficiencia en la generación de ítems.

La aplicación del *prompt* se realizó utilizando modelos de lenguaje de gran escala (LLM). Durante el proceso, se documentaron las interacciones entre las personas investigadoras y la IAG, registrando los ajustes realizados tras cada iteración. La secuencia permitió observar cómo el diseño de *prompts* estructurados facilita la coherencia temática de las respuestas, evita redundancias y mejora la pertinencia semántica de los ítems generados.

Este estudio aporta evidencia de que el diseño de instrumentos con IAG requiere un enfoque teórico-metodológico, donde la inteligencia humana y la generativa se complementan, porque la calidad de los ítems depende tanto de la precisión del prompt como del juicio experto que define el constructo, controla sesgos y verifica la validez de las respuestas. La IAG actúa como un medio



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

de amplificación cognitiva, no de sustitución, que optimiza tiempo, consistencia y alcance analítico, pero solo cuando opera bajo criterios teóricos explícitos y supervisión humana. La IAG actúa como un medio de amplificación cognitiva, no de sustitución, que optimiza tiempo, consistencia y alcance analítico, sin prescindir de la validación humana. Este hallazgo coincide con Flores Contrera (2024), quien advierte que los recursos generados por IAG son útiles en tareas estructuradas, pero limitados en contextos que exigen pensamiento crítico o comprensión situada.

La integración entre el modelo CIPP, la teoría de diseño de instrumentos y la ingeniería de *prompts* constituye una propuesta metodológica emergente en evaluación curricular. Si bien los resultados son prometedores, futuras investigaciones deberían validar estadísticamente los instrumentos generados y explorar la invariancia entre grupos y contextos, avanzando hacia una práctica evaluativa más justa, reproducible y tecnológicamente mediada.

## 5. Conclusiones

Este estudio contribuye a la consolidación de un enfoque metodológico que combina teoría educativa, fundamentos psicométricos y herramientas de IAG para el diseño de instrumentos de evaluación curricular. Su principal innovación radica en haber traducido la estructura del modelo CIPP y la teoría psicométrica de la validez y confiabilidad de los instrumentos educativos en una secuencia de *prompts* que permiten operacionalizar criterios de validez, coherencia y pertinencia curricular. Este enfoque integra la rigurosidad clásica de la evaluación con las posibilidades emergentes de la inteligencia artificial generativa, fortaleciendo la argumentación de validez desde un enfoque unificado (Messick, 1995).

Los resultados mostraron que el diseño de *prompts* estructurados y teóricamente fundamentados posibilita la generación de ítems congruentes con las dimensiones y criterios del modelo CIPP. El proceso iterativo de ajuste confirmó que la precisión semántica, la delimitación del rol y la claridad en los componentes [Instrucciones específicas de construcción] y [Formato] reducen los sesgos de respuesta y fortalecen la validez del producto generado. Estos hallazgos evidencian que la IAG, cuando es guiada por un marco teórico sólido y con conocimiento metodológico, puede producir resultados válidos, consistentes y ajustados a los objetivos de evaluación.

Desde una perspectiva crítica y ética, se reafirma que la IAG debe entenderse como una herramienta de co-creación y no de sustitución. Su aporte reside en potenciar la eficiencia y precisión en los procesos de diseño, siempre bajo supervisión humana y dentro de límites epistemológicos claramente definidos. La comprensión profunda de las teorías que sustentan la evaluación curricular permite al evaluador controlar el grado de creatividad del modelo, minimizar respuestas espurias y evitar distorsiones conceptuales.

La IAG se consolida como un agente metodológico complementario, capaz de optimizar el proceso de construcción de instrumentos cuando se utiliza con criterio ético, control semántico y rigor teórico. El desafío futuro radica en fortalecer los mecanismos de validación y transparencia, garantizando que la colaboración entre inteligencia humana y generativa contribuya a una evaluación educativa más justa, confiable y contextualizada.



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

**Agradecimiento a los revisores**

La Revista «La Universidad» agradece a los siguientes revisores por su evaluación y sugerencias en este artículo:

**Mtra. Liseth Guadalupe Oviedo de Artero**
Investigadora y docente de la Universidad Francisco Gavidia.
l.g.oviedo.guevara@gmail.com
**Mtro. Bladimir Antonio Olivar Miranda**
Coordinador de investigación del Instituto Tecnológico de Chalatenango.
bladimir.olivar@itcha.edu.sv

Sus aportes fueron fundamentales para mejorar la calidad y rigor de esta investigación.

## 6. Referencias

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Ahmady, S., Kohan, N., Mirmoghtadaie, Z. S., Hamidi, H., Divshali, B. S., & Rakhshani, T. (2023). Designing and psychometric analysis of an instrument to assess learning process in a virtual environment. *Smart Learning Environments, 10*, 35. https://doi.org/10.1186/s40561-023-00254-w

Aránguiz, K. (2020). *Propuesta de modelo de evaluación curricular para carreras y programas de pregrado de la Universidad Católica de la Santísima Concepción* [Tesis de maestría, Universidad del Desarrollo]. Repositorio UDD. https://repositorio.udd.cl/items/34ddca7d-60c8-46df-ac23-f50ae8d6e209

Arias, M. R. M., Lloreda, M. J. H., y Lloreda, M. V. H. (2014). *Psicometría*. Alianza Editorial.

Aziz, S., et al. (2018). Implementation of CIPP model for quality evaluation at school level: A case study. *Journal of Education and Educational Development, 5*(1), 189–206. https://files.eric.ed.gov/fulltext/EJ1180614.pdf

Carrillo-Cayllahua, J., Cóndor-Salvatierra, E., Oré-Rojas, J., y Gonzales-Castro, A. (2023). *Evaluación curricular de una carrera profesional en educación superior universitaria*. Inudi Perú. https://doi.org/10.35622/inudi.b.123

Cely, M., y Quiñones, A. (2022). Revisión sistemática de las características de evaluación curricular en programas académicos de pregrado a través del método PRISMA-NMA. *Revista Electrónica Calidad en la Educación Superior, 13*(2), 150–174. https://doi.org/10.22458/caes.v13i2.4415

Chamorro, D., y Borjas, M. (2020). *Investigación evaluativa curricular: Un camino a la transformación del aula*. Universidad del Norte. https://manglar.uninorte.edu.co/bitstream/handle/10584/9252/9789587892185%20eInvestigacion%20evaluativa%20curricular.pdf

Contreras, M. (2021). *Modelo integral para la evaluación curricular: De las variables a los instrumentos*. UNAM. https://www.zaragoza.unam.mx/wp-content/Portal2015/publicaciones/libros/csociales/Modelo_integral.pdf

Flores Contrera, C. J. (2024). La evaluación educativa en la era de la inteligencia artificial: Cambios de paradigmas. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades, 5*(1), 1579–1591. https://doi.org/10.56712/latam.v5i1.1694

Gregor, S., y Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly, 37*(2), 337–356.

Hernández-León, N., y Rodríguez-Conde, M. J. (2023). Inteligencia artificial aplicada a la educación y la evaluación educativa en la universidad. *Revista de Educación a Distancia (RED), 23*(71). https://doi.org/10.6018/red

Inciarte, A., y Canquiz, L. (2001). Análisis de la consistencia interna del currículo. *Informe de Investigaciones Educativas, 15*(1–2), 79–90. https://www.researchgate.net/publication/237265696_Analisis_de_la_consistencia_interna_del_curriculo

Karatas, H., y Fer, S. (2011). CIPP evaluation model scale: Development, reliability and validity. *Procedia - Social and Behavioral Sciences, 15*, 592–599. https://doi.org/10.1016/j.sbspro.2011.03.146

Korzynski, P., Mazurek, G., Krzypkowska, P., y Kurasinski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. *Entrepreneurial Business and Economics Review, 11*(3), 25–37. https://doi.org/10.15678/EBER.2023.110302

Lee, D., y Palmer, E. (2025). Prompt engineering in higher education: A systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education, 22*, 7. https://doi.org/10.1186/s41239-025-00503-7

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship, 49*, 102720. https://doi.org/10.1016/j.acalib.2023.102720

López, R., Valdez, A., y Martínez, J. (2020). Validez de constructo y confiabilidad de un instrumento para evaluar la docencia en educación superior. *Revista de Evaluación Educativa, 12*(3), 45–63.

Maaz, S., Palaganas, J. C., Palaganas, G., y Bajwa, M. (2025). A guide to prompt design: Foundations and applications for healthcare simulationists. *Frontiers in Medicine, 11*, 1504532. https://doi.org/10.3389/fmed.2024.1504532



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797

Méndez-Mantuano, M. O., Morán, M. Y. O., Mayorga, I. I. C., Valdez, A. Y. L., Rosado, Á. R. H., y Robles, D. V. A. (2024). La evaluación académica en la era de la inteligencia artificial (IA). *South Florida Journal of Development, 5*(1), 119–148. https://doi.org/10.46932/sfjdv5n1-010

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education & Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Mora, A. (2004). La evaluación educativa: Concepto, períodos y modelos. *Actualidades Investigativas en Educación, 4*(2). https://www.redalyc.org/pdf/447/44740211.pdf

Pizano, G. (2014). Modelos de evaluación curricular. *Revista Investigación Educativa, 4*(6), 15–22. https://revistasinvestigacion.unmsm.edu.pe/index.php/educa/article/view/7640/6649

Ramos-Armijos, D. F., Ramos-Armijos, D. G., Tapia-Puga, V. M., y Tapia-Puga, L. I. (2023). Explorando las fronteras: La aplicación de inteligencia artificial en la evaluación educativa. *Ciencia Latina Revista Científica Multidisciplinar, 7*(6), 5657–5667. https://doi.org/10.37811/cl_rcm.v7i6.9108

Reeves, T. C. (2006). Design research from a technology perspective. In J. van den Akker, K. Gravemeijer, S. McKenney, y N. Nieveen (Eds.), *An introduction to educational design research* (pp. 52–66). Netherlands Institute for Curriculum Development (SLO).

Schulhoff, S., Ilie, M., Balepur, N., et al. (2024). *The prompt report: A systematic survey of prompting techniques*. University of Maryland. https://arxiv.org/abs/2406.06608

Stufflebeam, D. L., y Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. Jossey-Bass.

Taureaux Díaz, N., Miralles Aguilera, E., Vicedo Tomey, A., y Díaz-Perera, G. (2016). Instrumento para la evaluación curricular de la función de administración en la carrera de Medicina. *Revista Habanera de Ciencias Médicas, 15*(5), 769–781.

Velásquez-Henao, J. D., Franco-Cardona, C. J., y Cadavid-Higuita, L. (2023). Prompt engineering: A methodology for optimizing interactions with AI-language models in the field of engineering. *Revista DYNA, 90*(230), 9–17. https://doi.org/10.15446/dyna.v90n230.111700

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., y Schmidt, D. C. (2023). *A prompt pattern catalog to enhance prompt engineering with ChatGPT*. Vanderbilt University. https://arxiv.org/abs/2302.11382



Universidad de El Salvador | Revista La Universidad | Vol. 7 | N. ° 2 , abril - junio 2026
ISSN: 0041-8242, E-ISSN: 3005-58-57, ISSN-L: 0041-8242 | ISNI: 0000 0001 2107 1797