

SSTSDS-Vol.4. N1. 102

Creación de guías y rúbricas de evaluación formativa y pedagógica para química orgánica a nivel de pregrado con IA para implementación en AulaWeb

Creation of formative and pedagogical assessment guides and rubrics for organic chemistry at the undergraduate level with AI for implementation in AulaWeb

Autores:

José Julio Barriga Jiménez
Universidad Popular del Cesar
Valledupar – Colombia

josebarrigajimenez@gmail.com
<https://orcid.org/0009-0007-5697-9069>

Ledys Alejandra Barliza Daza
Corporación para el Fomento Empresarial de Colombia – CORFECOL
Valledupar – Colombia

barlizadazaledys@gmail.com
<https://orcid.org/0009-0001-9543-0702>

Alcides Antonio Torregroza Mozo
Universidad Popular del Cesar
Valledupar – Colombia

alcidestorregroza@unicesar.edu.co
<https://orcid.org/0000-0003-1703-5368>

Jessica Andrea Plazas Moreno
Universidad Popular del Cesar
Valledupar – Colombia

jplazas@unicesar.edu.co
<https://orcid.org/0009-0002-9891-6422>

Javier Sair Molina Daza
Universidad Popular del Cesar
Valledupar – Colombia

jaric37@gmail.com
<https://orcid.org/0009-0006-1527-1622>

Autor de correspondencia: José Julio Barriga Jiménez, josebarrigajimenez@mail.com

Enviado: 13-05-2026
Aceptado: 03-08-2026

Revisado: 28-06-2026
Publicado: 05-08-2026



Cómo citar este artículo:

Barriga Jiménez, J. J., Barliza Daza, L. A., Torregroza Mozo, A. A., Plazas Moreno, J. A., & Molina Daza, J. S. (2026). Creación de guías y rúbricas de evaluación formativa y pedagógica para química orgánica a nivel de pregrado con IA para implementación en AulaWeb. *Sage Sphere of Technology, Sciences, Discoveries And Society*, 4(1), 1-21. <https://doi.org/10.63688/pbqd5p77>

© 2026; Los autores. Este es un artículo en acceso abierto, distribuido bajo los términos de una licencia Creative Commons (<https://creativecommons.org/licenses/by/4.0>) que permite el uso, distribución y reproducción en cualquier medio, siempre que la obra original sea correctamente citada.



RESUMEN

La evaluación de la química orgánica exige valorar no solo la respuesta final, sino también la comprensión conceptual, la representación molecular, el razonamiento mecanístico y la capacidad para corregir errores. El objetivo de esta revisión fue analizar la evidencia sobre evaluación formativa, rúbricas e inteligencia artificial aplicada a la educación química, con el propósito de establecer una propuesta de guías y rúbricas para estudiantes de pregrado en AulaWeb. Se realizó una revisión sistemática con síntesis cualitativa, reportada conforme a PRISMA 2020. Se evaluaron 22 publicaciones a texto completo; 18 respondieron directamente las preguntas de revisión y se incorporaron en la matriz de análisis, mientras que cuatro se utilizaron exclusivamente como apoyo teórico o metodológico. Los estudios se agruparon en cuatro categorías: evaluación formativa y rúbricas, inteligencia artificial en educación superior, inteligencia artificial en ciencias y química, y aplicaciones en química orgánica. Los hallazgos muestran que las rúbricas son más útiles cuando contienen criterios observables, niveles diferenciados, retroalimentación orientada a la acción y oportunidades de reelaboración. La inteligencia artificial puede apoyar la generación de actividades, descriptores y comentarios preliminares, pero también produce explicaciones plausibles con errores químicos. Se propone un modelo de implementación en AulaWeb basado en orientación, desarrollo, autoevaluación, retroalimentación y nueva entrega. Se concluye que la inteligencia artificial debe operar como asistencia pedagógica bajo validación disciplinar y supervisión docente.

Palabras clave: evaluación formativa, rúbricas, inteligencia artificial, química orgánica, educación superior.

ABSTRACT

Organic chemistry assessment must evaluate not only final answers but also conceptual understanding, molecular representation, mechanistic reasoning, and the ability to correct errors. This review analyzed evidence on formative assessment, rubrics, and artificial intelligence applied to chemistry education to establish a proposal for undergraduate guides and rubrics in AulaWeb. A systematic review with qualitative synthesis was conducted and reported according to PRISMA 2020. Twenty-two full-text publications were assessed; 18 directly answered the review questions and were incorporated into the analysis matrix, while four were used only as theoretical or methodological support. The studies were grouped into four categories: formative assessment and rubrics, artificial intelligence in higher education, artificial intelligence in science and chemistry, and applications in organic chemistry. Findings indicate that rubrics are most useful when they include observable criteria, differentiated levels, action-oriented feedback, and opportunities for revision. Artificial intelligence can support the generation of activities, descriptors, and preliminary comments, but it may also produce plausible explanations containing chemical errors. An AulaWeb implementation model based on orientation, development, self-assessment, feedback, and resubmission is proposed. Artificial intelligence should therefore operate as pedagogical assistance under disciplinary validation and teacher supervision.

Keywords: formative assessment, rubrics, artificial intelligence, organic chemistry, higher education.



1. INTRODUCCIÓN

El aprendizaje de la química orgánica representa uno de los componentes más exigentes de la formación científica universitaria. Los estudiantes deben interpretar estructuras bidimensionales y tridimensionales, reconocer grupos funcionales, relacionar propiedades con la estructura, anticipar la reactividad y explicar mecanismos mediante el movimiento de electrones. Estas tareas requieren integración conceptual y no pueden reducirse a la memorización de nombres, reactivos o productos. En consecuencia, la evaluación debe recuperar evidencias sobre el proceso de razonamiento y no limitarse a comprobar si la respuesta final coincide con una clave preestablecida.

La evaluación formativa ofrece un marco apropiado para atender esta necesidad. Black y Wiliam (2009) explican que una evaluación adquiere carácter formativo cuando la evidencia obtenida se utiliza para modificar la enseñanza y favorecer el progreso del estudiante. Desde esta perspectiva, el error constituye una fuente de información: una carga formal incorrecta, una flecha curva mal ubicada o la elección de un intermediario poco estable permiten identificar el punto en el que se interrumpe el razonamiento. El propósito del docente no consiste únicamente en marcar la equivocación, sino en proporcionar orientaciones que permitan comprenderla y superarla.

La retroalimentación es central en este proceso. Nicol y Macfarlane-Dick (2006) relacionan la evaluación formativa con la autorregulación, debido a que el estudiante necesita comparar su producción con criterios de calidad, reconocer la distancia entre el desempeño actual y el esperado, y planificar acciones de mejora. Carless y Boud (2018) añaden que la utilidad del feedback depende también de la alfabetización del estudiante para interpretarlo y emplearlo. En química orgánica, una observación como “mecanismo incorrecto” aporta poca información; en cambio, indicar que la flecha debe comenzar en un par electrónico disponible y terminar en el centro electrofílico orienta una revisión concreta.

Los entornos virtuales amplían las posibilidades de organizar ciclos de evaluación, comentario y nueva entrega. Gikandi et al. (2011) identificaron que la evaluación formativa en línea puede favorecer la participación y la reflexión cuando existe coherencia entre resultados de aprendizaje, tareas, criterios y retroalimentación. Esta articulación resulta pertinente para AulaWeb, donde las guías pueden organizar instrucciones, recursos,



evidencias, momentos de autoevaluación y criterios de calificación dentro de una misma secuencia pedagógica.

Las rúbricas hacen visibles los criterios y los niveles de desempeño. Panadero y Jonsson (2013) señalan que estos instrumentos pueden mejorar la transparencia, la autoevaluación y la autorregulación cuando se utilizan durante el proceso y no solamente al final. Brookhart y Chen (2015) destacan que una rúbrica formativa debe describir cualidades observables del trabajo, evitando niveles contruidos únicamente con adjetivos como excelente, bueno o regular. En química orgánica, los descriptores deben distinguir entre exactitud conceptual, representación molecular, razonamiento mecanístico, argumentación, uso de evidencia y capacidad de revisión.

La expansión de la inteligencia artificial generativa introduce nuevas posibilidades y desafíos. Zawacki-Richter et al. (2019) identificaron aplicaciones de inteligencia artificial en educación superior relacionadas con tutoría, predicción, evaluación y personalización, pero advirtieron una presencia limitada de perspectivas pedagógicas y docentes. Crompton y Burke (2023) confirmaron el crecimiento del campo y la diversidad de funciones atribuidas a estas tecnologías. Bond et al. (2024) señalaron, además, la necesidad de incrementar el rigor, la transparencia, la ética y la colaboración interdisciplinaria.

En evaluación, la inteligencia artificial puede generar preguntas, proponer niveles de desempeño, producir ejemplos y elaborar comentarios iniciales. Sin embargo, su fluidez verbal puede ocultar errores. Farrokhnia et al. (2024) describen esta dualidad: los modelos generativos ofrecen rapidez y personalización, pero también pueden producir información inexacta, sesgos y dependencia. Por ello, la calidad aparente de una respuesta no debe confundirse con exactitud científica.

Las investigaciones en química confirman esta precaución. Clark (2023) encontró que el desempeño de un chatbot variaba según el tipo de pregunta y disminuía ante tareas que requerían representaciones y razonamiento químico. Cooper (2023) observó potencial para apoyar la educación científica, aunque insistió en la verificación del contenido. En química orgánica, Watts et al. (2023) compararon respuestas estudiantiles y producciones generadas por inteligencia artificial, mientras que Yik y Dood (2024) mostraron que una explicación mecanística puede parecer sofisticada y contener inexactitudes relevantes.



El problema de investigación surge de la necesidad de integrar estas dos líneas: una evaluación formativa disciplinarmente sólida y un uso de inteligencia artificial que no sustituya el juicio pedagógico. Por tanto, el objetivo fue analizar la evidencia sobre evaluación formativa, rúbricas e inteligencia artificial aplicada a la enseñanza universitaria de química, con el propósito de establecer criterios para diseñar guías y rúbricas de química orgánica e implementarlas en AulaWeb. Las preguntas de revisión fueron: qué principios formativos deben incorporarse; qué dimensiones disciplinares deben evaluarse; qué funciones puede cumplir la inteligencia artificial; y qué condiciones de supervisión, trazabilidad y validación son necesarias.

2. METODOLOGÍA

Diseño de la revisión

Se desarrolló una revisión sistemática de literatura con alcance descriptivo e interpretativo y síntesis cualitativa. Debido a la heterogeneidad de los diseños, objetivos y resultados de las publicaciones, no se realizó metaanálisis. El reporte del proceso se organizó conforme a PRISMA 2020, utilizado como guía para transparentar la identificación, selección, elegibilidad e inclusión de los documentos (Page et al., 2021).

Estrategia de búsqueda y elegibilidad

La búsqueda temática se orientó hacia publicaciones sobre evaluación formativa, rúbricas, inteligencia artificial en educación superior, educación química y química orgánica. Se emplearon combinaciones en español e inglés, entre ellas: “formative assessment” AND “higher education”; “scoring rubrics” AND “formative assessment”; “rubrics” AND “chemistry education”; “artificial intelligence” AND “higher education assessment”; “generative artificial intelligence” AND “chemistry education”; “ChatGPT” AND “organic chemistry”; y “artificial intelligence” AND “reaction mechanisms”.

Para construir el corpus se evaluaron 52 publicaciones únicas a texto completo. Después de aplicar los criterios de inclusión y exclusión, se descartaron 30 documentos por no responder directamente a las preguntas de investigación, no centrarse en educación superior, presentar un enfoque exclusivamente tecnológico o no ofrecer resultados transferibles al diseño de instrumentos para química orgánica. Asimismo, cuatro publicaciones no se incorporaron en la síntesis cualitativa: tres se utilizaron únicamente como apoyo teórico general y una como



guía metodológica para el reporte PRISMA 2020. En consecuencia, el corpus final quedó conformado por 18 estudios, incluidos en la matriz de caracterización y en el análisis cualitativo.

Tabla 1

Criterios de inclusión y exclusión

Dimensión	Criterios de inclusión	Criterios de exclusión
Tipo de documento	Artículos científicos y revisiones académicas con datos bibliográficos verificables.	Editoriales, opiniones, publicidad o documentos sin sustento académico.
Pertinencia	Evaluación formativa, rúbricas, IA educativa, enseñanza de ciencias, química o química orgánica.	Trabajos sin relación con evaluación, educación superior o enseñanza de química.
Nivel educativo	Educación superior o resultados transferibles al pregrado.	Estudios exclusivamente escolares sin aplicabilidad universitaria.
Aporte	Resultados, modelos, riesgos o criterios útiles para diseñar guías y rúbricas.	Desarrollo computacional sin análisis pedagógico o disciplinar.
Disponibilidad	Texto completo o información suficiente para la extracción.	Registros incompletos o imposibles de verificar.

Nota. Los criterios fueron definidos según la pertinencia temática, el nivel educativo, el tipo de documento, la disponibilidad del texto completo y la relación de las publicaciones con los objetivos de la revisión. Elaboración propia.

Selección y análisis de la información

Las cuatro publicaciones no incorporadas en la síntesis fueron Black y Wiliam (2009), Nicol y Macfarlane-Dick (2006), Carless y Boud (2018) y Page et al. (2021). Las tres primeras se conservaron para fundamentar evaluación formativa, autorregulación y alfabetización en retroalimentación; la cuarta sustentó el reporte PRISMA. Por tanto, no fueron descartadas del manuscrito, sino separadas del corpus de resultados porque no respondían de manera

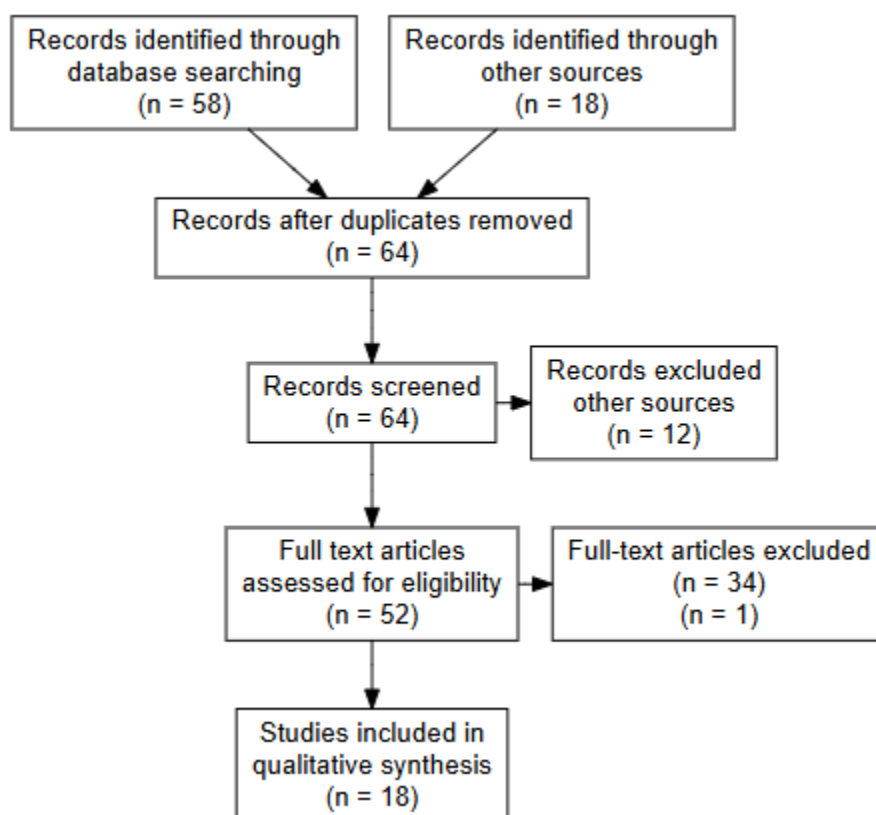


directa a las preguntas específicas sobre diseño de rúbricas, inteligencia artificial o educación química.

Para extraer la información se construyó una matriz con autoría, año, campo de aplicación, tipo de contribución, hallazgos y utilidad para el diseño de guías y rúbricas. Posteriormente se aplicó codificación temática. Las unidades de significado se agruparon en cuatro categorías: evaluación formativa y rúbricas; inteligencia artificial en educación superior; inteligencia artificial en ciencias y química; y aplicaciones en química orgánica. Finalmente se compararon convergencias, limitaciones y posibilidades de implementación en AulaWeb.

Figura 1.

Flujo de selección de publicaciones conforme a PRISMA 2020



Nota. El diagrama presenta el proceso de identificación, cribado, evaluación de elegibilidad e inclusión de los estudios. De los 76 registros identificados inicialmente, 18 cumplieron los criterios establecidos y fueron incorporados en la síntesis cualitativa. Elaboración propia a partir de Page et al. (2021).

Criterios de consistencia y calidad del análisis



Para fortalecer la consistencia de la síntesis, cada publicación fue examinada según su correspondencia con las preguntas de revisión, la claridad de su propósito, el contexto educativo descrito y la utilidad de sus resultados para el diseño de instrumentos. No se construyó una puntuación numérica de calidad, porque el corpus integró revisiones, estudios exploratorios, intervenciones educativas y aplicaciones de aula con diseños difícilmente comparables. En su lugar, se aplicó una valoración cualitativa orientada a reconocer la solidez del aporte, sus limitaciones y el grado de transferencia a la enseñanza de química orgánica. La síntesis evitó atribuir a los estudios conclusiones que no hubieran sido examinadas en sus respectivos contextos. Los resultados sobre chatbots, evaluación automatizada o secuencias didácticas se interpretaron como evidencia de posibilidades y riesgos, no como prueba de eficacia universal. Asimismo, la propuesta para AulaWeb se formuló como un modelo pedagógico derivado de la literatura y no como una intervención ya validada. Esta distinción permite separar los hallazgos documentales de las recomendaciones que deberán comprobarse posteriormente mediante aplicación con docentes y estudiantes.

3. RESULTADOS

Caracterización del corpus

Los 18 estudios incluidos se distribuyeron en cuatro categorías. La primera reunió investigaciones sobre evaluación formativa en línea, diseño de rúbricas y valoración de habilidades de laboratorio. La segunda agrupó revisiones sobre aplicaciones, beneficios, riesgos y principios éticos de la inteligencia artificial en educación superior. La tercera comprendió estudios de inteligencia artificial en ciencias, química y evaluación automatizada. La cuarta se concentró en química orgánica, mecanismos de reacción, prompts y actividades de laboratorio.

Tabla 2

Estudios incluidos en la síntesis cualitativa

Referencia	Área	Tipo	Aporte principal
Gikandi et al. (2011)	Evaluación formativa virtual	Revisión	Articulación entre tareas, criterios, interacción y retroalimentación en educación superior.
Panadero y Jonsson (2013)	Rúbricas formativas	Revisión	Las rúbricas apoyan transparencia, autorregulación y mejora



			cuando se usan durante la tarea.
Brookhart y Chen (2015)	Calidad de rúbricas	Revisión	Los descriptores deben ser específicos, observables y claramente diferenciados.
Harwood et al. (2020)	Laboratorio de química	Aplicación educativa	Criterios observables para valorar preparación, procedimiento, seguridad y desempeño práctico.
Zawacki-Richter et al. (2019)	IA en educación superior	Revisión sistemática	Predominio tecnológico y necesidad de mayor participación pedagógica y docente.
Crompton y Burke (2023)	IA en educación superior	Revisión	Aplicaciones en tutoría, evaluación, predicción, personalización y apoyo al aprendizaje.
Bond et al. (2024)	IA en educación superior	Metarrevisión	Necesidad de rigor, ética, transparencia y colaboración interdisciplinaria.
Farrokhnia et al. (2024)	ChatGPT en educación	Análisis SWOT	Rapidez y personalización frente a errores, sesgos, opacidad y dependencia.
Clark (2023)	Química general	Estudio evaluativo	El rendimiento del chatbot cambia según el formato y la demanda de razonamiento químico.
Cooper (2023)	Educación científica	Estudio exploratorio	Potencial de apoyo acompañado de verificación científica y alfabetización crítica.
Costa et al. (2024)	Secuencia didáctica de química	Estudio comparativo	La IA genera propuestas organizadas, pero puede ser genérica y poco sensible al contexto.
Du et al. (2024)	Evaluación automatizada	Estudio aplicado	Los modelos lingüísticos apoyan la revisión preliminar de informes con criterios definidos.
Erümit y Özdemir Sarılioğlu (2025)	IA en ciencias y química	Revisión sistemática	Aplicaciones en personalización, tutoría, evaluación y generación de contenidos.
Tassoti (2024)	Prompts en química	Estudio educativo	La orientación sobre prompts mejora el uso reflexivo de la inteligencia artificial.
Okoth (2025)	Química orgánica	Intervención formativa	Las actividades digitales breves favorecen participación y comprobación de conceptos.



Ruff et al. (2024)	Laboratorio y química verde	Aplicación educativa	La IA funciona como punto de partida, pero exige contraste y corrección.
Watts et al. (2023)	Escritura sobre mecanismos	Estudio comparativo	Las respuestas de IA difieren de las estudiantiles en razonamiento y uso de conceptos.
Yik y Dood (2024)	Mecanismos orgánicos	Análisis de respuestas	Explicaciones sofisticadas pueden contener errores conceptuales y mecanísticos.

Nota. La tabla presenta los 18 estudios incorporados directamente en la síntesis cualitativa. Las cuatro referencias de apoyo teórico y metodológico no se incluyeron en esta matriz. Elaboración propia.

Tabla 3

Categorías de análisis y aplicación al diseño de instrumentos

Categoría	Estudios	Hallazgo integrado	Aplicación en AulaWeb
Evaluación formativa y rúbricas	4	Criterios explícitos, feedback orientado a acciones, autoevaluación y reelaboración.	Publicar guía y rúbrica desde el inicio; habilitar revisión y nueva entrega.
IA en educación superior	4	La personalización y rapidez requieren ética, transparencia y participación docente.	Usar IA como apoyo, mantener validación humana y declarar su utilización.
IA en ciencias y química	6	Genera actividades y comentarios, pero puede fallar ante simbolismo y razonamiento complejo.	Validar bancos de preguntas, ejemplos y retroalimentaciones antes de publicarlos.
Aplicaciones en química orgánica	4	Las explicaciones pueden ser sofisticadas sin ser químicamente correctas.	Evaluar exactitud, mecanismo, fuentes, prompts y correcciones realizadas.

Nota. Las categorías se establecieron mediante la agrupación temática de los principales hallazgos identificados en los 18 estudios incluidos en la revisión. Elaboración propia.

Principios para las guías de evaluación formativa

La síntesis mostró que una guía formativa debe comunicar con anticipación el resultado de aprendizaje, el producto esperado, los recursos permitidos y la secuencia de trabajo. También debe presentar la rúbrica antes de la entrega, incorporar una actividad de autoevaluación y ofrecer una oportunidad real de corrección. Gikandi et al. (2011) respaldan la integración entre actividad, criterios y feedback en entornos virtuales. Panadero y Jonsson (2013) y Brookhart y Chen (2015) coinciden en que la rúbrica adquiere valor formativo cuando orienta la producción y no se utiliza únicamente para justificar una nota.



En química orgánica, los criterios deben responder a la naturaleza de la disciplina. La exactitud de una estructura depende de enlaces, cargas, pares electrónicos y estereoquímica; la validez de un mecanismo depende del movimiento electrónico, la estabilidad de intermediarios y la coherencia entre reactivos, condiciones y productos. Harwood et al. (2020) muestran que las rúbricas también pueden valorar habilidades observables de laboratorio, como preparación, manipulación, seguridad, registro y disposición de residuos.

Tabla 4
Estructura propuesta para las guías formativas de química orgánica

Componente	Contenido mínimo	Finalidad formativa
Resultado de aprendizaje	Competencia observable y contenido específico.	Delimitar qué evidencia debe producirse.
Situación o problema	Caso, reacción, mecanismo, síntesis o práctica.	Activar aplicación y razonamiento.
Instrucciones	Pasos, formato, recursos y tiempo.	Reducir ambigüedades operativas.
Uso de IA	Herramientas permitidas, registro de prompts y reglas de citación.	Promover transparencia y análisis crítico.
Rúbrica	Criterios, niveles, ponderación y ejemplos.	Comunicar la calidad esperada.
Autoevaluación	Preguntas de control y revisión previa.	Favorecer autorregulación.
Retroalimentación	Comentarios específicos vinculados con criterios.	Orientar acciones de mejora.
Reelaboración	Nueva entrega y justificación de cambios.	Convertir el error en aprendizaje.

Nota. La estructura propuesta integra los principios de evaluación formativa, autorregulación, retroalimentación y uso crítico de inteligencia artificial identificados en la literatura revisada. Elaboración propia.

Funciones y riesgos de la inteligencia artificial

Los estudios sobre inteligencia artificial en educación superior evidenciaron funciones de tutoría, generación de contenido, evaluación preliminar y personalización. Zawacki-Richter et al. (2019), Crompton y Burke (2023) y Bond et al. (2024) coinciden en que el desarrollo tecnológico debe acompañarse de participación docente, criterios éticos y transparencia. Farrokhnia et al. (2024) advierten que las fortalezas de rapidez y accesibilidad conviven con riesgos de errores, sesgos, opacidad y dependencia.



En educación química, la inteligencia artificial mostró utilidad para crear preguntas, secuencias didácticas y comentarios preliminares. Clark (2023) y Cooper (2023) señalan que las respuestas deben verificarse, especialmente cuando incluyen símbolos, representaciones o razonamientos complejos. Costa et al. (2024) observaron que una secuencia generada por inteligencia artificial puede ser ordenada, pero también genérica, repetitiva o poco sensible al contexto. Du et al. (2024) muestran que los modelos lingüísticos pueden apoyar la evaluación de informes, aunque la decisión final depende de criterios claros y revisión académica.

Erümit y Özdemir Sarıalioğlu (2025) sintetizan aplicaciones de inteligencia artificial en ciencias y química relacionadas con personalización, tutoría, evaluación y generación de recursos. Tassoti (2024) destaca que la calidad de la interacción mejora cuando los estudiantes reciben orientación para construir prompts y revisar las respuestas. Esto justifica incorporar en la rúbrica una dimensión de uso crítico de inteligencia artificial, basada en trazabilidad, verificación, corrección y declaración del proceso.

Las investigaciones específicas en química orgánica reforzaron la necesidad de supervisión. Okoth (2025) mostró el valor de actividades digitales breves como evaluación formativa, siempre que se articulen con tareas de mayor complejidad. Ruff et al. (2024) emplearon inteligencia artificial como punto de partida para actividades de laboratorio y química verde, pero identificaron respuestas vagas o incorrectas que debían ser contrastadas. Watts et al. (2023) y Yik y Dood (2024) demostraron que los textos generados pueden presentar lenguaje sofisticado sin garantizar exactitud conceptual o mecanística.

Rúbrica analítica propuesta

A partir de estos hallazgos se diseñó una propuesta de implementación en AulaWeb. La secuencia contempla cinco momentos: orientación, desarrollo, autoevaluación, retroalimentación y reelaboración. La inteligencia artificial se ubica como apoyo en la creación inicial de actividades, ejemplos y comentarios, mientras que la validación de contenidos, la ponderación de criterios y la calificación final permanecen bajo responsabilidad docente.

Tabla 5

Rúbrica para actividades de química orgánica con apoyo de inteligencia artificial



Criterio	Avanzado	Satisfactorio	Básico	Insuficiente
Exactitud conceptual	Aplica correctamente conceptos y relaciones químicas sin errores relevantes.	Comprensión correcta con imprecisiones menores.	Comprensión parcial con confusiones que afectan parte de la respuesta.	Errores conceptuales que invalidan el resultado.
Representación molecular	Estructuras, enlaces, cargas y estereoquímica son completas y correctas.	Representaciones comprensibles con errores menores.	Omisiones o errores que dificultan la interpretación.	Representaciones incoherentes o químicamente imposibles.
Razonamiento mecanístico	Justifica etapas, movimiento electrónico e intermediarios.	Explica las etapas principales con justificación suficiente.	Describe etapas sin sustentar decisiones esenciales.	Presenta productos o flechas sin razonamiento coherente.
Resolución y argumentación	Selecciona estrategia pertinente y la sustenta con principios y evidencia.	Resuelve la mayor parte y ofrece justificación adecuada.	Necesita apoyo y presenta argumentos incompletos.	No identifica estrategia o formula afirmaciones sin sustento.
Uso crítico de IA	Registra prompts, verifica, corrige y explica decisiones con fuentes.	Declara el uso y realiza verificación suficiente.	Usa IA con comprobación superficial o trazabilidad incompleta.	Copia respuestas sin análisis, verificación ni declaración.
Reflexión y mejora	Identifica el origen de los errores y justifica cada corrección.	Corrige los errores principales a partir del feedback.	Realiza correcciones parciales sin explicación suficiente.	No utiliza la retroalimentación para mejorar.

Nota. Los criterios y niveles de desempeño fueron formulados a partir de los hallazgos sobre evaluación formativa, rúbricas disciplinares, razonamiento químico y uso responsable de inteligencia artificial. Elaboración propia.

Modelo de implementación en AulaWeb

La implementación se organiza en cinco momentos articulados. En orientación, el docente publica la guía, la rúbrica, ejemplos y reglas de uso de inteligencia artificial. Durante el desarrollo, el estudiante resuelve la actividad y conserva evidencias del proceso. En autoevaluación, compara su producción con los criterios. En retroalimentación, recibe comentarios automáticos previamente validados y observaciones docentes. Finalmente, en reelaboración, entrega una nueva versión y explica los cambios. Las herramientas concretas pueden adaptarse a las funciones disponibles en AulaWeb, manteniendo esta secuencia pedagógica.

Componentes mínimos de una guía aplicable en AulaWeb

El análisis permitió precisar que la guía no debe limitarse a presentar instrucciones operativas. Debe comenzar con un resultado de aprendizaje formulado como desempeño



observable, por ejemplo, explicar un mecanismo de sustitución mediante el movimiento electrónico y la estabilidad de los intermediarios. A continuación, debe plantear una situación que obligue a aplicar el conocimiento y no solamente a reproducir una definición. La evidencia solicitada puede incluir estructuras, flechas curvas, explicación escrita, comparación entre alternativas y una breve reflexión sobre los errores identificados.

También es necesario distinguir entre recursos autorizados y productos esperados. Cuando se permita el uso de inteligencia artificial, la guía debe indicar si puede utilizarse para explorar ideas, revisar redacción, generar ejemplos o contrastar respuestas. Debe solicitarse el registro de los prompts relevantes y de las modificaciones realizadas. Esta condición evita que la declaración de uso se convierta en un trámite superficial y permite observar cómo el estudiante evaluó la información generada.

La retroalimentación debe vincularse de manera directa con los criterios. Para exactitud conceptual, el comentario debe señalar el concepto afectado; para representación molecular, debe identificar enlaces, cargas o estereoquímica; y para razonamiento mecanístico, debe localizar la etapa en la que se rompe la coherencia electrónica. Los comentarios generales pueden acompañar la evaluación, pero no sustituyen las orientaciones específicas que permiten actuar sobre el error.

Finalmente, la nueva entrega debe solicitar una explicación breve de los cambios. Esta comparación entre versiones convierte la corrección en una evidencia adicional y permite valorar si el estudiante comprendió el feedback. En AulaWeb, la implementación puede realizarse con una primera entrega, un formulario de autoevaluación, la devolución de la rúbrica y una segunda entrega acompañada de una nota reflexiva. El valor formativo depende de que exista tiempo suficiente para aplicar las observaciones antes de cerrar la actividad.

4. DISCUSIÓN

Los resultados permiten sostener que la innovación no consiste en añadir una herramienta de inteligencia artificial a una evaluación tradicional, sino en reorganizar la evaluación alrededor de evidencias de razonamiento, retroalimentación y mejora. Una guía formativa debe explicar qué se aprenderá, cómo se demostrará y qué acciones podrán realizarse después de recibir comentarios. Este enfoque desplaza la atención desde la respuesta correcta aislada hacia el proceso mediante el cual el estudiante interpreta, justifica y corrige.



La rúbrica propuesta responde a una dificultad frecuente: los instrumentos generales pueden valorar presentación o cumplimiento, pero no discriminan la calidad del razonamiento químico. Los criterios de exactitud conceptual, representación molecular y mecanismo permiten observar dimensiones diferentes. Un estudiante podría dibujar correctamente el producto y, al mismo tiempo, justificarlo con un mecanismo incoherente. Una calificación global ocultaría esta diferencia; una rúbrica analítica permite identificarla y convertirla en retroalimentación.

La inteligencia artificial puede reducir parte de la carga asociada con la elaboración inicial de actividades y comentarios, pero no elimina la necesidad de competencia disciplinar. Los hallazgos de Clark (2023), Watts et al. (2023) y Yik y Dood (2024) muestran que los modelos pueden construir respuestas convincentes con errores. En consecuencia, un sistema que automatice la retroalimentación sin validación podría reforzar conceptos incorrectos. La supervisión docente no es un complemento opcional, sino una condición de calidad.

La propuesta introduce el uso crítico de inteligencia artificial como objeto de evaluación. En lugar de limitarse a prohibir o autorizar la herramienta, se solicita al estudiante documentar prompts, comparar respuestas, verificar afirmaciones y justificar cambios. Este procedimiento transforma la interacción con la tecnología en evidencia de aprendizaje. También permite distinguir entre dependencia y aprovechamiento reflexivo.

La implementación en AulaWeb debe considerar que la retroalimentación automática funciona mejor ante errores frecuentes y claramente definidos. Por ejemplo, puede orientar sobre nomenclatura, reconocimiento de grupos funcionales o conservación de cargas. En cambio, los problemas de síntesis, los mecanismos alternativos y las decisiones experimentales requieren interpretación contextual. Por ello, se recomienda combinar bancos de comentarios validados con observaciones personalizadas.

El modelo propuesto también favorece la coherencia entre evaluación individual y colaboración. La autoevaluación permite revisar el trabajo antes de entregarlo; la coevaluación puede utilizarse con ejemplos anónimos; y la retroalimentación docente orienta la versión final. Este ciclo es compatible con la finalidad formativa descrita en la literatura y evita que la rúbrica sea percibida como una lista aplicada únicamente al cierre.

Una limitación de la revisión es el tamaño reducido del corpus específico sobre química orgánica e inteligencia artificial. La mayor parte de las publicaciones recientes son



exploratorias y analizan versiones concretas de modelos que pueden cambiar con rapidez. Asimismo, esta investigación propone instrumentos a partir de literatura, pero no evalúa todavía su desempeño con estudiantes. Futuras investigaciones deberían aplicar la propuesta, someter las rúbricas a juicio de expertos, medir acuerdo entre evaluadores y analizar si la retroalimentación mejora la calidad de mecanismos, explicaciones y procedimientos experimentales.

Validez, integridad académica y sostenibilidad de la propuesta

La validez de la evaluación depende de que la actividad mida aprendizaje de química orgánica y no solamente habilidad para producir texto con una herramienta. Por ello, las tareas deben requerir decisiones que puedan explicarse, representaciones que deban comprobarse y relaciones entre estructura y reactividad. Una respuesta extensa obtenida mediante inteligencia artificial no constituye por sí sola evidencia de comprensión. La rúbrica debe otorgar mayor importancia a la coherencia química, la argumentación y la verificación que a la fluidez o longitud del texto.

Desde la integridad académica, la trazabilidad ofrece una alternativa más educativa que la prohibición absoluta. Solicitar prompts, fuentes consultadas y correcciones permite reconocer el aporte de la herramienta y la responsabilidad intelectual del estudiante. No obstante, esta política debe comunicarse antes de la actividad y aplicarse de manera consistente. Las reglas ambiguas pueden generar desigualdad entre quienes declaran el uso y quienes lo ocultan.

La sostenibilidad de la propuesta exige mantener un banco institucional de guías, criterios, ejemplos y comentarios revisados. Los docentes pueden reutilizar y adaptar estos recursos, pero deben verificar cada actualización cuando cambien el contenido, el nivel del curso o la herramienta de inteligencia artificial. La capacitación debe incluir diseño de rúbricas, detección de errores químicos, protección de datos y criterios de transparencia. De esta manera, AulaWeb funcionaría como espacio de organización pedagógica y no solamente como repositorio de tareas.

5. CONCLUSIONES

La evaluación formativa en química orgánica debe orientarse a la obtención de evidencias sobre el proceso de razonamiento del estudiante, en lugar de limitarse únicamente a la



valoración de la respuesta final. En este sentido, las guías de aprendizaje deben articular de manera coherente los resultados de aprendizaje, las actividades propuestas, los criterios de evaluación, la retroalimentación y las oportunidades de revisión y reelaboración.

Las rúbricas, por su parte, deben incorporar criterios disciplinares claros, observables y medibles, entre los que se incluyen la exactitud conceptual, la correcta representación de estructuras moleculares, el razonamiento mecanístico, la resolución de problemas, la capacidad de argumentación, la aplicación de normas de seguridad experimental y la reflexión crítica sobre los errores cometidos.

La inteligencia artificial puede emplearse como una herramienta de apoyo para la elaboración preliminar de actividades, ejemplos, descriptores de desempeño y comentarios de retroalimentación. No obstante, los contenidos generados deben someterse a un proceso de revisión y validación docente, debido a que estos sistemas pueden producir respuestas lingüísticamente coherentes, pero conceptual o químicamente incorrectas.

El uso de inteligencia artificial por parte de los estudiantes debe evaluarse mediante criterios relacionados con la trazabilidad del proceso, la pertinencia y calidad de los prompts utilizados, la verificación de la información mediante fuentes confiables, la identificación de errores y la justificación de las modificaciones efectuadas sobre las respuestas generadas.

La implementación de estas herramientas en AulaWeb debe estructurarse como un ciclo continuo de orientación, desarrollo de actividades, autoevaluación, retroalimentación, corrección y nueva entrega. Dentro de este proceso, el docente debe conservar la responsabilidad de validar la exactitud disciplinar de los contenidos y adoptar la decisión evaluativa final.

REFERENCIAS BIBLIOGRÁFICAS

- Black, P., y Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
<https://doi.org/10.1007/s11092-008-9068-5>
- Bond, M., Khosravi, H., de Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., y Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International*



Journal of Educational Technology in Higher Education, 21, Article 4.
<https://doi.org/10.1186/s41239-023-00436-z>

Brookhart, S. M., y Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, 67(3), 343–368.
<https://doi.org/10.1080/00131911.2014.929565>

Carless, D., y Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>

Clark, T. M. (2023). Investigating the use of an artificial intelligence chatbot with general chemistry exam questions. *Journal of Chemical Education*, 100(5), 1905–1916.
<https://doi.org/10.1021/acs.jchemed.3c00027>

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32(3), 444–452. <https://doi.org/10.1007/s10956-023-10039-y>

Costa, G. G., Nascimento Junior, W. J. D., Mombelli, M. N., y Giroto Junior, G. (2024). Revisiting a teaching sequence on the topic of electrolysis: A comparative study with the use of artificial intelligence. *Journal of Chemical Education*, 101(8), 3255–3263.
<https://doi.org/10.1021/acs.jchemed.4c00247>

Crompton, H., y Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, Article 22. <https://doi.org/10.1186/s41239-023-00392-8>

Du, H., Jia, Q., Gehringer, E., y Wang, X. (2024). Harnessing large language models to auto-evaluate the student project reports. *Computers and Education Artificial Intelligence*, 7, Article 100268. <https://doi.org/10.1016/j.caeai.2024.100268>

Erümit, A. K., y Özdemir Sarıalioğlu, R. (2025). Artificial intelligence in science and chemistry education: A systematic review. *Discover Education*, 4, Article 178.
<https://doi.org/10.1007/s44217-025-00622-3>

Farrokhnia, M., Banihashem, S. K., Noroozi, O., y Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460–474.
<https://doi.org/10.1080/14703297.2023.2195846>



- Gikandi, J. W., Morrow, D., y Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers and Education*, 57(4), 2333–2351. <https://doi.org/10.1016/j.compedu.2011.06.004>
- Harwood, C. J., Hewett, S., y Towns, M. H. (2020). Rubrics for assessing hands-on laboratory skills. *Journal of Chemical Education*, 97(7), 2033–2035. <https://doi.org/10.1021/acs.jchemed.0c00200>
- Nicol, D. J., y Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Okoth, R. (2025). Using Kahoot! as a formative assessment tool in undergraduate organic chemistry lecture courses. *Journal of Chemical Education*, 102(7), 3004–3009. <https://doi.org/10.1021/acs.jchemed.4c01538>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, 10, Article 89. <https://doi.org/10.1186/s13643-021-01626-4>
- Panadero, E., y Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Ruff, E. F., Franz, J. L., y West, J. K. (2024). Using ChatGPT for method development and green chemistry education in upper-level laboratory courses. *Journal of Chemical Education*, 101(8), 3224–3232. <https://doi.org/10.1021/acs.jchemed.4c00193>
- Tassoti, S. (2024). Assessment of students use of generative artificial intelligence: Prompting strategies and prompt engineering in chemistry education. *Journal of Chemical Education*, 101(6), 2475–2482. <https://doi.org/10.1021/acs.jchemed.4c00212>
- Watts, F. M., Dood, A. J., Shultz, G. V., y Rodriguez, J. M. G. (2023). Comparing student and generative artificial intelligence chatbot responses to organic chemistry writing-to-learn assignments. *Journal of Chemical Education*, 100(10), 3806–3817. <https://doi.org/10.1021/acs.jchemed.3c00664>



Yik, B. J., y Dood, A. J. (2024). ChatGPT convincingly explains organic chemistry reaction mechanisms slightly inaccurately with high levels of explanation sophistication. *Journal of Chemical Education*, 101(5), 1836–1846. <https://doi.org/10.1021/acs.jchemed.4c00235>

Zawacki-Richter, O., Marín, V. I., Bond, M., y Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>

CONFLICTO DE INTERESES

Ninguno.

FINANCIAMIENTO

No aplica.

CONTRIBUCIÓN DE AUTORÍA

Conceptualización: José Julio Barriga Jiménez, Ledys Alejandra Barliza Daza y Alcides Antonio Torregroza Mozo.

Metodología: José Julio Barriga Jiménez, Jessica Andrea Plazas Moreno y Javier Sair Molina Daza.

Investigación: José Julio Barriga Jiménez, Ledys Alejandra Barliza Daza, Alcides Antonio Torregroza Mozo, Jessica Andrea Plazas Moreno y Javier Sair Molina Daza.

Redacción – borrador original: José Julio Barriga Jiménez, Ledys Alejandra Barliza Daza y Alcides Antonio Torregroza Mozo.

Redacción – revisión y edición: José Julio Barriga Jiménez, Ledys Alejandra Barliza Daza, Alcides Antonio Torregroza Mozo, Jessica Andrea Plazas Moreno y Javier Sair Molina Daza.

