



Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), septiembre-octubre 2026,
Volumen 10, Número 5.

https://doi.org/10.37811/cl_rcm.v10i5

EVALUACIÓN HÍBRIDA Y RETROALIMENTACIÓN FORMATIVA: UN MODELO HUMAN IN THE LOOP PARA LA DIDÁCTICA UNIVERSITARIA

**HYBRID ASSESSMENT AND FORMATIVE FEEDBACK: A
HUMAN-IN-THE-LOOP MODEL FOR UNIVERSITY
TEACHING**

Rosalynn Tejada Auccacusi

Universidad Nacional de San Antonio Abad del Cusco, Perú

Federico Ubaldo Fernandez Sutta

Universidad Nacional de San Antonio Abad del Cusco, Perú

Delfín Richard Mendoza Quispe

Universidad Nacional de San Antonio Abad del Cusco, Perú

Jose Angel Yañez Tuero

Universidad Nacional de San Antonio Abad del Cusco, Perú

Evaluación Híbrida y Retroalimentación Formativa: Un Modelo Human in the Loop para la Didáctica Universitaria

Rosalynn Tejada Auccacusi¹

rosalynn.tejada@unsaac.edu.pe

<https://orcid.org/0000-0001-6588-9515>

Universidad Nacional de San Antonio Abad
del Cusco
Perú

Federico Ubaldo Fernandez Sutta

federico.fernandez@unsaac.edu.pe

<https://orcid.org/0000-0002-3453-6589>

Universidad Nacional de San Antonio Abad
del Cusco
Perú

Delfin Richard Mendoza Quispe

delfin.mendoza@unsaac.edu.pe

<https://orcid.org/0009-0003-3915-1412>

Universidad Nacional de San Antonio Abad
del Cusco
Perú

Jose Angel Yañez Tuero

joseyanez2298@gmail.com

<https://orcid.org/0009-0001-3494-4443>

Universidad Nacional de San Antonio Abad
del Cusco
Perú

RESUMEN

La incorporación de sistemas de inteligencia artificial generativa a los procesos evaluativos de la educación superior ha desplazado el debate desde la factibilidad técnica hacia la definición del lugar que ocupa el juicio profesional docente en el ciclo de retroalimentación. Este artículo tuvo por objetivo construir un modelo conceptual de evaluación híbrida con retroalimentación formativa bajo el principio human in the loop, aplicable a la didáctica universitaria latinoamericana, que explicita las funciones evaluativas delegables a sistemas automatizados y aquellas irreducibles a la intervención del profesional de la educación. Se desarrolló una revisión documental sistematizada de treinta y cuatro publicaciones científicas arbitradas, recuperadas de repositorios indexados y verificadas individualmente mediante resolución de identificadores digitales de objeto, sometidas a análisis de contenido categorial con codificación axial. Los hallazgos se organizaron en tres categorías: la retroalimentación como cuello de botella estructural de la evaluación híbrida, las capacidades y límites documentados de la retroalimentación generada por sistemas, y la configuración del bucle de intervención docente. El modelo resultante, denominado HITL-DU, distribuye el ciclo evaluativo en cinco momentos —encuadre, generación asistida, curaduría docente, devolución dialógica y metaevaluación— y asigna a cada uno un rol profesional específico, un indicador de calidad y un riesgo asociado a la delegación total. Se concluye que la contribución pedagógica de la herramienta tecnológica no reside en su capacidad generativa sino en la arquitectura de decisión que el docente impone sobre ella, lo que reposiciona la competencia digital docente como competencia de gobierno del juicio evaluativo y no como destreza instrumental.

Palabras Clave: evaluación formativa, retroalimentación, inteligencia artificial, educación superior, didáctica universitaria

¹ Autor principal.

Correspondencia: rosalynn.tejada@unsaac.edu.pe

Hybrid Assessment and Formative Feedback: A Human-In-The-Loop Model for University Teaching

ABSTRACT

The incorporation of generative artificial intelligence systems into higher education assessment processes has shifted the debate from technical feasibility toward defining the place occupied by professional teacher judgement within the feedback cycle. This article aimed to construct a conceptual model of hybrid assessment with formative feedback under the human-in-the-loop principle, applicable to Latin American university teaching, making explicit which assessment functions may be delegated to automated systems and which remain irreducible to the intervention of the education professional. A systematized documentary review of thirty-four peer-reviewed scientific publications was developed, retrieved from indexed repositories and individually verified through digital object identifier resolution, and subjected to categorical content analysis with axial coding. Findings were organized into three categories: feedback as the structural bottleneck of hybrid assessment, the documented capacities and limits of system-generated feedback, and the configuration of the teacher intervention loop. The resulting model, designated HITL-DU, distributes the assessment cycle across five moments —framing, assisted generation, teacher curation, dialogic delivery and meta-assessment— assigning to each a specific professional role, a quality indicator and a risk associated with full delegation. It is concluded that the pedagogical contribution of the technological tool lies not in its generative capacity but in the decision architecture the teacher imposes upon it, repositioning teacher digital competence as competence in governing assessment judgement rather than as instrumental skill.

Keywords: formative assessment, feedback, artificial intelligence, higher education, university teaching

Artículo recibido 10 agosto 2026

Aceptado para publicación: 15 septiembre 2026



Avaliação Híbrida e Feedback Formativo: Um Modelo Humano no Ciclo para Didática Universitária

RESUMO

A incorporação de sistemas de inteligência artificial generativa aos processos avaliativos do ensino superior deslocou o debate da viabilidade técnica para a definição do lugar ocupado pelo juízo profissional docente no ciclo de retroalimentação. Este artigo teve como objetivo construir um modelo conceitual de avaliação híbrida com retroalimentação formativa sob o princípio *human in the loop*, aplicável à didática universitária latino-americana, explicitando as funções avaliativas delegáveis a sistemas automatizados e aquelas irredutíveis à intervenção do profissional da educação. Desenvolveu-se uma revisão documental sistematizada de trinta e quatro publicações científicas arbitradas, recuperadas de repositórios indexados e verificadas individualmente mediante resolução de identificadores digitais de objeto, submetidas a análise de conteúdo categorial com codificação axial. Os achados foram organizados em três categorias: a retroalimentação como gargalo estrutural da avaliação híbrida, as capacidades e limites documentados da retroalimentação gerada por sistemas, e a configuração do laço de intervenção docente. O modelo resultante, denominado HITL-DU, distribui o ciclo avaliativo em cinco momentos —enquadramento, geração assistida, curadoria docente, devolução dialógica e metaavaliação— e atribui a cada um um papel profissional específico, um indicador de qualidade e um risco associado à delegação total. Conclui-se que a contribuição pedagógica da ferramenta tecnológica não reside em sua capacidade generativa, mas na arquitetura de decisão que o docente impõe sobre ela.

Palavras-chave: avaliação formativa, retroalimentação, inteligência artificial, ensino superior, didática universitária



INTRODUCCIÓN

La evaluación de los aprendizajes constituye, desde hace décadas, el componente de la práctica universitaria donde con mayor nitidez se manifiestan las tensiones entre las aspiraciones formativas del discurso pedagógico y las restricciones materiales del ejercicio docente. El desplazamiento conceptual que condujo desde una evaluación concebida como certificación terminal hacia una evaluación entendida como dispositivo de regulación del aprendizaje fue documentado tempranamente por la investigación educativa, que estableció que las prácticas evaluativas orientadas a informar la enseñanza producen ganancias de aprendizaje sustantivas, particularmente entre los estudiantes con menor rendimiento previo (Black y Wiliam, 1998). Ese hallazgo fundacional instaló una agenda que continúa vigente y que sitúa a la retroalimentación —entendida como la información que permite al estudiante reducir la distancia entre su desempeño actual y el desempeño esperado— en el centro de la eficacia didáctica (Hattie y Timperley, 2007). Sin embargo, tres décadas de producción científica sobre el tema han revelado una paradoja persistente: aquello que la evidencia identifica como la intervención de mayor impacto potencial es también aquello que las condiciones reales de trabajo docente hacen más difícil de sostener con calidad y frecuencia suficientes.

La reformulación de la retroalimentación como proceso y no como producto profundizó esa tensión. La propuesta de abandonar los modelos centrados en la transmisión de información correctiva para adoptar diseños en los que el estudiante interviene activamente en la elaboración y el uso de la devolución recibida desplazó el problema desde la producción del mensaje hacia la creación de condiciones para su apropiación (Boud y Molloy, 2013). Esta perspectiva encontró desarrollo posterior en la noción de alfabetización en retroalimentación, que designa el conjunto de capacidades, disposiciones y saberes que permiten al estudiante comprender, valorar y utilizar la información evaluativa para mejorar su trabajo (Carless y Boud, 2018). La consolidación de este marco implicó un reconocimiento incómodo: ningún incremento en la cantidad de retroalimentación produce, por sí solo, mejoras en el aprendizaje si no existe una arquitectura didáctica que asegure su procesamiento. La calidad de la devolución dejó de definirse por sus atributos internos y pasó a definirse por sus efectos en la actividad posterior del estudiante, lo que exige del docente una intervención que excede ampliamente la corrección.



En paralelo a esta maduración conceptual, la organización material de la enseñanza superior experimentó una transformación acelerada. La masificación del acceso, la diversificación de los perfiles de ingreso y la consolidación de modalidades que combinan presencialidad y virtualidad modificaron las coordenadas en las que se ejerce la función evaluativa. La experiencia de virtualización forzada durante la emergencia sanitaria global expuso con crudeza la fragilidad de los dispositivos evaluativos concebidos para el aula presencial y desencadenó una reflexión sobre la naturaleza misma de la evaluación en entornos mediados, donde la observación directa del proceso se vuelve indirecta y la inferencia sobre el aprendizaje depende de evidencias construidas deliberadamente (Bennasar-García y Estrada, 2021). La salida de esa emergencia no restauró el estado anterior: consolidó, en cambio, configuraciones híbridas estables cuya valoración por parte de docentes y estudiantes resulta heterogénea y está condicionada por la disciplina, la infraestructura disponible y la experiencia previa de cada actor (Pattier y Ferreira, 2023). La hibridación dejó de ser una contingencia y se transformó en una condición estructural de la didáctica universitaria contemporánea, con implicaciones directas sobre la gestión del tiempo docente y la trazabilidad de las evidencias de aprendizaje.

El estudio de experiencias híbridas concretas ha mostrado que la simple superposición de canales no produce integración didáctica. El análisis de una propuesta de enseñanza en vivo que articula participantes presenciales y remotos en tiempo real reveló que la calidad del dispositivo depende de decisiones de diseño deliberadas sobre la distribución de la atención, la construcción de la participación y la articulación de las evidencias generadas en cada canal, y no de la disponibilidad tecnológica (Ruiz Domínguez y Maggio, 2025). Esta constatación es relevante para el presente trabajo porque anticipa el argumento central: en los escenarios híbridos, el valor pedagógico no reside en la mediación técnica sino en la intencionalidad que la organiza. La misma lógica resulta aplicable, como se argumentará, a la incorporación de sistemas inteligentes al ciclo evaluativo.

Sobre ese escenario irrumpió, con una velocidad de adopción sin precedentes en la historia reciente de la tecnología educativa, la inteligencia artificial generativa. Su disponibilidad masiva y su bajo umbral de acceso modificaron simultáneamente las prácticas de estudio, las condiciones de producción de evidencias evaluables y las posibilidades de generación automatizada de devoluciones. La literatura especializada ha documentado tanto el potencial transformador de estos sistemas como los riesgos



asociados a su incorporación acrítica, entre los que destacan la reproducción de sesgos presentes en los datos de entrenamiento, la opacidad de los criterios que subyacen a sus salidas y la erosión de la autonomía intelectual de los actores educativos (Cordón García, 2023; Del Valle, 2025). El mapeo bibliométrico del campo confirma que la producción sobre ética e inteligencia artificial en educación ha crecido de manera sostenida, aunque con un desbalance persistente entre la formulación de principios normativos y la disponibilidad de dispositivos operativos que permitan traducirlos en decisiones didácticas concretas (Santana-Soriano, 2025).

En el terreno específico de la retroalimentación, la evidencia disponible es prometedora pero incompleta. Las revisiones sistemáticas sobre retroalimentación automatizada generada por sistemas generativos en educación superior reportan beneficios consistentes en términos de inmediatez, cobertura y disponibilidad, junto con limitaciones recurrentes relativas a la precisión disciplinar, la contextualización curricular y la profundidad del diagnóstico (Lee y Moore, 2024). Los estudios sobre el compromiso estudiantil con devoluciones generadas por sistemas de evaluación automatizada de la escritura muestran patrones de apropiación selectiva: los estudiantes incorporan con facilidad las sugerencias de superficie y desatienden las observaciones de orden estructural o argumentativo, lo que reproduce en un nuevo soporte la dificultad histórica de la retroalimentación de proceso (Yeung, 2025). En la misma dirección, la investigación sobre la interacción de estudiantes con devoluciones conversacionales generadas por modelos de lenguaje advierte que el compromiso productivo depende críticamente del nivel previo de alfabetización en retroalimentación del estudiante, de modo que la herramienta amplifica las diferencias existentes en lugar de compensarlas (Zhan y Yan, 2026).

Este conjunto de hallazgos configura el problema que motiva la presente investigación. La discusión pública sobre inteligencia artificial en evaluación universitaria ha oscilado entre dos posiciones igualmente insuficientes: una posición sustitutiva, que concibe el sistema como reemplazo del trabajo docente en tareas consideradas rutinarias, y una posición prohibicionista, que lo concibe como amenaza a la integridad académica y busca su exclusión del proceso. Ambas comparten un supuesto erróneo: que la relación entre docente y sistema es de suma cero. La perspectiva *human in the loop* —literalmente, humano en el bucle, entendida como la configuración de sistemas automatizados en los que la intervención humana se incorpora deliberadamente como componente estructural del ciclo de decisión



y no como control externo posterior— ofrece un marco alternativo. La revisión del campo muestra que su aplicación en educación se ha concentrado en el diseño de sistemas adaptativos y en la anotación de datos, con un desarrollo comparativamente escaso en el dominio de la evaluación formativa y una definición imprecisa del tipo de intervención humana requerida en cada punto del ciclo (Memarian y Doleck, 2024).

La formulación de este marco en clave educativa ha sido planteada con claridad conceptual: el cierre del bucle no constituye un requisito técnico sino una exigencia pedagógica, porque las decisiones evaluativas comprometen trayectorias formativas y responsabilidades profesionales que no admiten delegación algorítmica (Ninaus y Sailer, 2022). La investigación reciente sobre analítica del aprendizaje ha reforzado esta posición al situar el fomento de la agencia humana como criterio de diseño de los sistemas inteligentes, en oposición a las arquitecturas que optimizan la eficiencia predictiva a costa de la capacidad de decisión de los actores (Viberg et al., 2025). Los estudios empíricos sobre percepciones estudiantiles de analíticas potenciadas por inteligencia artificial en procesos de retroalimentación confirman que la legitimidad de la devolución ante el estudiante depende de la visibilidad de la mediación docente: cuando el estudiante percibe que la devolución fue producida sin intervención profesional, disminuye su disposición a utilizarla (Jin et al., 2025).

En el contexto latinoamericano, este problema adquiere rasgos específicos. Las condiciones de ejercicio de la docencia universitaria en la región —elevadas relaciones estudiante-docente, alta proporción de contratación por horas, heterogeneidad de la formación pedagógica del profesorado y desigualdad en el acceso a infraestructura— configuran un escenario en el que la presión por automatizar es considerablemente mayor y las salvaguardas institucionales considerablemente menores. Los estudios sobre competencia digital docente en educación superior han señalado que el concepto ha sido operacionalizado predominantemente como dominio instrumental de herramientas, con un tratamiento marginal de las dimensiones críticas, éticas y de toma de decisiones (Kroff Trujillo y Bellemans Fagalde, 2026). Las indagaciones sobre percepción docente del uso de inteligencia artificial en la región muestran actitudes ambivalentes, con reconocimiento del potencial de alivio de carga y preocupación simultánea por la pérdida de control sobre criterios evaluativos (Valenzuela-Velasco, 2026). El análisis de las competencias requeridas al docente investigador en este nuevo escenario converge en la necesidad de



desarrollar capacidades de discernimiento que exceden el manejo operativo de las herramientas (Cribilleros-Shigihara et al., 2025).

Existe, en consecuencia, una brecha identificable entre la abundante producción sobre los efectos de la retroalimentación automatizada y la escasez de modelos didácticos que especifiquen, con nivel de detalle operativo, qué funciones evaluativas pueden delegarse a un sistema, cuáles requieren intervención docente ineludible y qué indicadores permiten verificar que esa intervención efectivamente ocurrió. Esa brecha es particularmente crítica para la didáctica universitaria latinoamericana, donde la adopción tecnológica suele preceder a la formulación de criterios pedagógicos y donde la ausencia de modelos explícitos favorece configuraciones de facto que desplazan el juicio profesional sin decisión institucional deliberada. En atención a lo expuesto, el objetivo de este artículo consiste en construir un modelo conceptual de evaluación híbrida con retroalimentación formativa bajo el principio human in the loop, aplicable a la didáctica universitaria, que explicita las funciones evaluativas delegables a sistemas automatizados y aquellas irreductibles a la intervención del profesional de la educación, junto con los indicadores que permiten verificar dicha intervención en cada momento del ciclo evaluativo.

METODOLOGÍA

Enfoque y tipo de investigación

La investigación se inscribe en el paradigma cualitativo interpretativo y adopta un diseño de revisión documental sistematizada con finalidad de construcción teórica. Esta definición requiere una precisión metodológica explícita: el estudio no constituye una revisión sistemática en sentido estricto, dado que no aplicó un protocolo de registro previo ni un procedimiento de cribado con evaluación independiente por pares revisores ni una síntesis cuantitativa de efectos. La denominación revisión documental sistematizada designa un procedimiento que conserva la sistematicidad en la definición de criterios, la trazabilidad de las fuentes y la explicitación de las decisiones analíticas, sin reclamar la exhaustividad ni la replicabilidad estricta propias del protocolo sistemático formal. Esta distinción se declara aquí de manera deliberada, en coherencia con el principio de honestidad metodológica que exige no atribuir a un diseño garantías epistémicas que no puede ofrecer.

El alcance es descriptivo-interpretativo con proyección propositiva. El estudio no busca establecer relaciones causales ni estimar magnitudes de efecto, sino identificar regularidades en la producción



científica disponible y organizarlas en un dispositivo conceptual de utilidad didáctica. La construcción del modelo se realizó por vía inferencial a partir de las categorías emergentes del análisis, y su estatuto es el de una propuesta teórica fundamentada en literatura, pendiente de validación empírica.

Corpus documental y criterios de selección

El universo documental estuvo constituido por publicaciones científicas arbitradas disponibles en repositorios de acceso abierto y bases de datos de indexación regional e internacional. La búsqueda se realizó mediante consultas bibliográficas estructuradas sobre la interfaz programática de Crossref, complementadas con exploración directa en portales editoriales de revistas indexadas en SciELO, Redalyc, Dialnet y el Directorio de Revistas de Acceso Abierto.

Los descriptores empleados combinaron términos en español y en inglés en cuatro ejes: evaluación formativa y retroalimentación; evaluación híbrida y modalidades mixtas en educación superior; inteligencia artificial generativa aplicada a evaluación y retroalimentación; y human in the loop, agencia humana y juicio profesional docente. La combinación de descriptores en ambos idiomas respondió al propósito de asegurar representación de la producción iberoamericana, habitualmente subrepresentada en las búsquedas monolingües en inglés.

Se establecieron como criterios de inclusión: (a) artículos publicados en revistas científicas con arbitraje; (b) ventana temporal 2021-2026, con admisión excepcional de obras seminales anteriores cuando constituyen referencia fundacional inexcusable del campo; (c) tratamiento sustantivo de al menos uno de los cuatro ejes descriptores; (d) disponibilidad de identificador digital de objeto verificable; y (e) texto completo accesible. Se excluyeron: (a) documentos sin arbitraje, incluidas actas de congreso no indexadas, materiales de divulgación y publicaciones preliminares no evaluadas; (b) trabajos centrados exclusivamente en niveles educativos no universitarios sin proyección transferible a la educación superior; (c) estudios sobre inteligencia artificial en educación sin vinculación con procesos evaluativos o de retroalimentación; y (d) fuentes cuya metadata bibliográfica no pudo verificarse.

El procedimiento de verificación constituyó un componente central del diseño. Cada referencia candidata fue sometida a resolución individual de su identificador digital de objeto contra el registro de la agencia de asignación, con extracción de metadatos de autoría, título, publicación, año, volumen, número y paginación. Las fuentes cuyo identificador no resolvió o cuyos metadatos resultaron



inconsistentes fueron descartadas y sustituidas. Este control responde al riesgo, documentado en la literatura reciente sobre integridad académica, de incorporación de referencias no verificables en trabajos elaborados con asistencia de sistemas generativos. El corpus final quedó conformado por treinta y cuatro documentos, de los cuales treinta corresponden al período 2021-2026 y cuatro a obras seminales del campo de la evaluación formativa.

Técnicas e instrumentos

La técnica de análisis empleada fue el análisis de contenido categorial de orientación inductiva. El instrumento consistió en una matriz de codificación construida ad hoc, organizada en cinco campos por documento: identificación bibliográfica, objeto y diseño del estudio, hallazgos relativos a la función del docente en el ciclo evaluativo, hallazgos relativos a las capacidades o límites de la mediación tecnológica, y unidad de sentido extraída en forma de proposición.

Procedimiento

El desarrollo se organizó en cuatro fases secuenciales. La primera fase, de identificación, comprendió la ejecución de las consultas bibliográficas estructuradas y la conformación de un conjunto preliminar de candidatas a partir de la revisión de títulos, resúmenes y palabras clave. La segunda fase, de depuración y verificación, aplicó los criterios de inclusión y exclusión y sometió cada fuente superviviente al procedimiento de resolución de identificadores descrito. La tercera fase, de codificación, consistió en la lectura analítica de los documentos incorporados y su volcado en la matriz, con identificación de unidades de sentido y asignación de códigos abiertos. La cuarta fase, de construcción, agrupó los códigos abiertos en categorías mediante codificación axial y derivó de esas categorías la arquitectura del modelo propuesto, verificando la trazabilidad de cada componente del modelo respecto de las unidades de sentido que lo sustentan.

Análisis de datos

El análisis siguió una lógica de saturación categorial. Los códigos abiertos se agruparon por afinidad semántica hasta la estabilización de tres categorías, que se presentan en el apartado de resultados. La validez del procedimiento se apoyó en tres criterios: trazabilidad, entendida como la posibilidad de reconstruir el recorrido desde cada afirmación del modelo hasta la fuente que la sustenta; suficiencia, entendida como la presencia de al menos dos fuentes independientes para cada componente categorial;



y contraste, entendido como la búsqueda deliberada de evidencia que tensionara las categorías en formación, con incorporación de hallazgos discrepantes en el análisis.

Consideraciones éticas y limitaciones

El estudio no involucró participantes humanos ni datos personales, por lo que no requirió dictamen de comité de ética. Se observaron los principios de integridad académica en la atribución de autoría, la representación fiel de los hallazgos de las fuentes y la abstención absoluta de fabricación de datos o referencias. Se reconocen tres limitaciones. Primero, la restricción idiomática a producción en español, inglés y portugués excluye literatura relevante en otras lenguas. Segundo, la ausencia de doble codificación independiente impide reportar índices de acuerdo entre codificadores, lo que constituye una limitación reconocida del diseño. Tercero, el modelo propuesto posee estatuto teórico y no ha sido sometido a validación empírica en aula, condición que se señala explícitamente como agenda de investigación derivada.

RESULTADOS

El análisis del corpus permitió la estabilización de tres categorías que organizan la presentación de los hallazgos. La primera caracteriza la retroalimentación como punto de estrangulamiento estructural de la evaluación en contextos híbridos. La segunda delimita el perímetro de capacidades y límites documentados de la retroalimentación generada por sistemas automatizados. La tercera describe la configuración del bucle de intervención docente y constituye el núcleo del modelo propuesto. La Tabla 1 presenta la composición del corpus según eje temático, procedencia regional y período de publicación.

Tabla 1 Composición del corpus documental según eje temático, procedencia y período

Eje temático	Fuentes (n)	Procedencia predominante	Período
Evaluación formativa y retroalimentación	11	Iberoamérica	1998-2024
Evaluación híbrida y modalidades mixtas	4	Iberoamérica	2021-2025
Inteligencia artificial y retroalimentación automatizada	10	Internacional	2022-2026
Human in the loop, agencia y juicio docente	9	Internacional e Iberoamérica	2022-2026
Total	34	—	1998-2026

Nota. Elaboración propia a partir del corpus verificado. La categoría procedencia refiere a la filiación institucional predominante de los equipos autorales, no al idioma de publicación.



Categoría 1: la retroalimentación como cuello de botella estructural de la evaluación híbrida

El corpus converge en la identificación de una asimetría persistente entre el reconocimiento del valor formativo de la retroalimentación y su materialización efectiva en la práctica universitaria. Los estudios que examinan la retroalimentación desde la perspectiva docente coinciden en señalar que las restricciones de tiempo, la magnitud de los grupos y la ausencia de formación específica operan como determinantes de la calidad de la devolución, con mayor peso que las concepciones pedagógicas declaradas por el profesorado (Muñoz Cortés et al., 2023; Parra Martínez et al., 2022). El diagnóstico se repite con notable estabilidad a través de contextos disciplinares y nacionales diversos, lo que sugiere un carácter estructural antes que situacional.

El análisis de las estrategias e instrumentos efectivamente utilizados revela un repertorio limitado, con predominio de modalidades escritas unidireccionales y baja incidencia de dispositivos dialógicos o de seguimiento longitudinal (Valdez Valdez et al., 2023). Los trabajos que han intentado sistematizar intervenciones de retroalimentación formativa documentan que la eficacia depende menos del contenido del mensaje que de la existencia de una oportunidad estructurada de reelaboración posterior por parte del estudiante, condición que rara vez está prevista en el diseño de los cursos (Gómez Reyes, 2023). Esta observación resulta convergente con la reformulación procesual de la retroalimentación y confirma que el problema no reside en la producción de información sino en su integración en la secuencia didáctica (Boud y Molloy, 2013).

La instrumentación de la calidad de la retroalimentación mediante escalas validadas ha permitido precisar sus dimensiones constitutivas y verificar que los docentes que declaran prácticas evaluativas formativas no necesariamente producen devoluciones de calidad cuando estas se miden con criterios operativos (Dávila Ramírez y Huertas Martínez, 2024). Este desacople entre discurso y práctica adquiere especial relevancia en escenarios híbridos, donde la evidencia de aprendizaje se encuentra distribuida entre canales con distinta capacidad de captura y la reconstrucción del proceso exige un esfuerzo de integración adicional (Bennasar-García y Estrada, 2021; Ruiz Domínguez y Maggio, 2025).

La articulación de este conjunto de hallazgos con los marcos de evaluación auténtica y evaluación por competencias permite precisar la naturaleza del estrangulamiento. Los enfoques que orientan la evaluación hacia desempeños contextualizados y productos complejos incrementan la demanda de juicio



profesional por unidad de estudiante evaluado, precisamente porque las evidencias generadas son menos estandarizables (Huerta Rosales y Vigoa Escobedo, 2024). La literatura sobre evaluación de competencias en educación superior ha señalado que la multiplicación de agentes evaluadores — autoevaluación, evaluación entre pares, evaluación docente— responde en parte a esta presión, aunque su implementación efectiva requiere condiciones de formación y andamiaje que a menudo no se cumplen (Bonnefoy Valdés, 2021). El resultado es un escenario en el que las orientaciones pedagógicamente más avanzadas son también las más costosas de sostener, lo que genera una presión estructural hacia la automatización.

Categoría 2: capacidades y límites documentados de la retroalimentación generada por sistemas

La segunda categoría organiza la evidencia disponible sobre el desempeño efectivo de los sistemas generativos en tareas de retroalimentación. La síntesis sistemática del campo reporta un patrón consistente: los sistemas alcanzan desempeños satisfactorios en la detección de errores de superficie, en la generación de devoluciones extensas y en la disponibilidad inmediata, y desempeños significativamente inferiores en la valoración de la pertinencia disciplinar, la calibración de la exigencia respecto del nivel formativo y la identificación de errores conceptuales de fondo (Lee y Moore, 2024). Los estudios de compromiso estudiantil con devoluciones automatizadas de la escritura confirman esta asimetría desde la perspectiva del receptor y documentan una apropiación selectiva orientada a las sugerencias de menor complejidad cognitiva (Yeung, 2025). La evidencia generada en contextos universitarios iberoamericanos resulta convergente: el uso de inteligencia artificial generativa para retroalimentar aprendizajes produce valoraciones estudiantiles favorables en términos de claridad y oportunidad, pero reclama del docente una función de contextualización curricular que los estudiantes identifican como condición de utilidad de la devolución recibida (Leiva-Guerrero et al., 2026).

La investigación comparativa entre valoración humana y valoración automatizada en evaluación de escritura académica ha encontrado niveles variables de consistencia y ha identificado sesgos sistemáticos en las salidas de los sistemas, con tendencias diferenciadas según características del texto y del autor (Cui, 2026). Este hallazgo resulta particularmente relevante para el contexto latinoamericano, donde la producción escrita estudiantil incorpora variedades lingüísticas regionales y convenciones retóricas que pueden ser penalizadas por sistemas entrenados predominantemente sobre corpus



anglófonos. La literatura sobre riesgos éticos de la inteligencia artificial en educación había anticipado esta problemática al advertir sobre la reproducción algorítmica de sesgos y la opacidad de los criterios subyacentes (Del Valle, 2025; Santana-Soriano, 2025).

Las experiencias de diseño que han incorporado deliberadamente componentes humanos ofrecen evidencia favorable a la configuración híbrida. El desarrollo de marcos de razonamiento mixto para la generación de rúbricas de evaluación muestra que la combinación de generación automatizada y validación humana produce instrumentos superiores a los obtenidos por cualquiera de los dos procedimientos por separado (Mendonça et al., 2026). En una dirección convergente, los sistemas de retroalimentación automatizada dirigidos al propio docente —orientados a informar su reflexión sobre la práctica antes que a sustituir su intervención— han mostrado efectos positivos sobre el cambio instruccional, lo que sugiere que el destinatario más productivo de la automatización puede no ser el estudiante sino el profesional que diseña la enseñanza (Jacobs et al., 2022).

Un último conjunto de hallazgos aborda la dimensión de legitimidad. Los estudios sobre percepción estudiantil de analíticas potenciadas por inteligencia artificial en procesos de retroalimentación documentan que la disposición del estudiante a utilizar la devolución depende de la visibilidad de la mediación docente, y que la atribución de la devolución a un sistema sin intervención profesional reduce su valor percibido con independencia de su calidad objetiva (Jin et al., 2025). Esta constatación introduce una dimensión frecuentemente ausente del debate técnico: la eficacia de la retroalimentación no es una propiedad del mensaje sino de la relación pedagógica en la que se inscribe. La evidencia sobre el compromiso con devoluciones conversacionales generadas por modelos de lenguaje refuerza esta lectura al mostrar que el aprovechamiento se concentra en los estudiantes con alfabetización en retroalimentación previamente desarrollada (Zhan y Yan, 2026; Woitt et al., 2025). La Tabla 2 sintetiza la distribución de funciones evaluativas resultante del análisis.



Tabla 2 Matriz de distribución de funciones evaluativas entre sistema automatizado e intervención docente

Función evaluativa	Aporte del sistema	Aporte irreducible del docente	Riesgo de delegación total
Definición de criterios	Propuesta de borrador de rúbrica	Validación curricular y ajuste al nivel formativo	Criterios descontextualizados
Detección de errores de superficie	Alta cobertura y velocidad	Jerarquización según relevancia formativa	Saturación de observaciones irrelevantes
Diagnóstico conceptual	Identificación parcial y variable	Interpretación del error como indicio de comprensión	Diagnóstico erróneo no detectado
Calibración de la exigencia	Ausente o inestable	Ajuste a la trayectoria del estudiante	Desaliento o complacencia
Formulación del mensaje	Redacción extensa e inmediata	Selección de tono, foco y oportunidad	Devolución genérica y desafectada
Cierre del ciclo	Registro de interacciones	Diseño de la oportunidad de reelaboración	Retroalimentación sin uso posterior

Nota. Elaboración propia a partir del análisis categorial del corpus. La columna de riesgo describe el efecto esperable cuando la función se transfiere íntegramente al sistema sin validación profesional.

Categoría 3: configuración del bucle y estructura del modelo HITL-DU

La tercera categoría integra los hallazgos anteriores en una propuesta de organización del ciclo evaluativo. La revisión del campo human in the loop en educación había señalado la imprecisión conceptual de la noción de intervención humana y la necesidad de especificar su naturaleza en cada punto del ciclo (Memarian y Doleck, 2024). El modelo aquí denominado HITL-DU —human in the loop para la didáctica universitaria— responde a esa necesidad mediante la distribución del ciclo en cinco momentos, cada uno con un rol profesional definido, un indicador de calidad verificable y un riesgo asociado a la delegación total.

El primer momento, de encuadre, sitúa la intervención docente antes de cualquier operación del sistema. Comprende la definición del objeto de evaluación, la explicitación de los criterios y la determinación del nivel de exigencia esperable. Este momento es el que otorga al sistema su marco de referencia y condiciona la pertinencia de todas sus salidas posteriores; su omisión constituye la falla más frecuente y la de consecuencias más extendidas, dado que ningún control ulterior puede corregir un encuadre ausente. El segundo momento, de generación asistida, delega en el sistema la producción de un borrador

de devolución sobre la evidencia estudiantil, operación en la que el aporte diferencial de la herramienta es máximo en términos de cobertura y velocidad.

El tercer momento, de curaduría docente, constituye el núcleo del modelo y aquel donde se juega la diferencia entre una configuración human in the loop efectiva y una supervisión meramente nominal.

La curaduría comprende cuatro operaciones diferenciadas: verificación de la corrección disciplinar de las afirmaciones del borrador; jerarquización de las observaciones según su relevancia formativa, con supresión de aquellas que saturan sin orientar; recontextualización respecto de la trayectoria del estudiante y del momento del curso; y ajuste del registro comunicativo. Esta operación se apoya en el juicio profesional entendido como capacidad de discernimiento situado, y su naturaleza es la de una decisión pedagógica, no la de un control de calidad técnico (Ninaus y Sailer, 2022; Viberg et al., 2025).

El cuarto momento, de devolución dialógica, restituye la retroalimentación a su condición de proceso relacional. Comprende la entrega de la devolución curada, la habilitación de un espacio de consulta y la previsión explícita de una oportunidad de reelaboración con consecuencias sobre la valoración final. La evidencia indica que este momento es el que determina la conversión de la devolución en aprendizaje efectivo y que su ausencia anula el valor de todo el ciclo previo (Carless y Boud, 2018; Tam, 2024). El quinto momento, de metaevaluación, incorpora un ciclo de segundo orden en el que el docente examina el desempeño agregado del sistema sobre el conjunto de devoluciones producidas, identifica patrones de error recurrentes y ajusta el encuadre para iteraciones siguientes, operación que convierte la experiencia acumulada en mejora del dispositivo. La Tabla 3 presenta la estructura completa del modelo.

Tabla 3 Estructura del modelo HITL-DU: momentos, rol docente, indicador de calidad y riesgo

Momento	Rol del profesional de la educación	Indicador de calidad verificable	Riesgo si se omite
1. Encuadre	Diseñador del marco de referencia	Existencia de rúbrica validada previa a la generación	Criterios implícitos e inauditables
2. Generación asistida	Operador con propósito definido	Registro de instrucción y versión del sistema	Salidas no trazables
3. Curaduría docente	Curador del juicio evaluativo	Proporción de observaciones modificadas o suprimidas	Supervisión nominal sin decisión

4.	Devolución Interlocutor pedagógico dialógica	Existencia de instancia de reelaboración con efecto	Devolución sin uso posterior
5.	Metaevaluación Auditor del dispositivo	Registro de ajustes al encuadre por ciclo	Perpetuación de errores sistemáticos

Nota. Elaboración propia. Los indicadores de calidad fueron formulados con criterio de verificabilidad institucional, de modo que su cumplimiento pueda auditarse mediante evidencias documentales del propio proceso.

Un rasgo transversal del modelo merece destacarse. La secuencia sitúa la intervención docente en los extremos del ciclo —encuadre inicial y metaevaluación final— y en su punto medio —curaduría—, reservando para el sistema únicamente el segmento de generación. Esta arquitectura invierte la configuración implícita en la mayoría de las adopciones espontáneas observadas en la literatura, donde la herramienta se incorpora al inicio del proceso y la intervención humana se reduce a una revisión final de las salidas producidas (Albarrán Martínez, 2026). La inversión no es un detalle de secuencia: determina si el docente gobierna el dispositivo o si el dispositivo gobierna la evaluación.

DISCUSIÓN

Los resultados permiten sostener una tesis central: la contribución pedagógica de la herramienta tecnológica en el ciclo de retroalimentación no reside en su capacidad generativa sino en la arquitectura de decisión que el profesional de la educación impone sobre ella. Esta formulación desplaza el eje del debate desde la pregunta por la calidad del sistema hacia la pregunta por la calidad del diseño didáctico que lo incorpora, y encuentra sustento convergente en tres líneas de evidencia analizadas.

La primera línea proviene de los estudios comparativos entre generación automatizada y validación humana. El hallazgo de que los marcos de razonamiento mixto producen instrumentos de evaluación superiores a los obtenidos por cualquiera de los procedimientos aislados admite una interpretación que excede la constatación de complementariedad (Mendonça et al., 2026). Lo que el diseño mixto aporta no es la suma de dos capacidades sino la introducción de un momento de deliberación sobre criterios que el procedimiento puramente automatizado elude y que el procedimiento puramente manual tiende a naturalizar. El sistema, al producir una propuesta explícita, obliga al docente a posicionarse frente a criterios que en la práctica habitual permanecen tácitos. En esta lectura, el valor didáctico de la herramienta es en parte mayéutico: fuerza la explicitación del juicio profesional antes que reemplazarlo.



La segunda línea proviene de la evidencia sobre legitimidad y compromiso estudiantil. La constatación de que la disposición del estudiante a utilizar una devolución depende de la visibilidad de la mediación docente resulta difícil de explicar desde una concepción informacional de la retroalimentación (Jin et al., 2025). Si la devolución fuera únicamente transferencia de información correctiva, su origen sería irrelevante mientras su contenido fuera preciso. El hecho de que no lo sea confirma la concepción relacional consolidada por la literatura sobre alfabetización en retroalimentación, según la cual la devolución opera dentro de un vínculo pedagógico que le confiere autoridad y que orienta la disposición del receptor a modificar su desempeño (Carless y Boud, 2018; Tam, 2024; Woitt et al., 2025). Este resultado tiene una consecuencia práctica de primer orden para el diseño institucional: la mediación docente no solo debe existir, debe ser visible para el estudiante, lo que sugiere la conveniencia de explicitar el procedimiento seguido antes que ocultar la participación del sistema.

La tercera línea proviene de la evidencia sobre apropiación selectiva y amplificación de desigualdades. Los hallazgos que documentan que el aprovechamiento de la retroalimentación automatizada se concentra en los estudiantes con mayor alfabetización previa introducen una advertencia decisiva para contextos de alta heterogeneidad de ingreso, característicos de la expansión del acceso a la educación superior latinoamericana (Yeung, 2025; Zhan y Yan, 2026). Un dispositivo que incrementa el volumen de devolución disponible sin intervenir sobre las condiciones de su apropiación no es neutral respecto de la equidad: produce beneficios diferenciales que agravan la brecha inicial. El cuarto momento del modelo propuesto —devolución dialógica con oportunidad estructurada de reelaboración— responde directamente a este riesgo y constituye, desde una perspectiva de justicia curricular, el componente menos prescindible del ciclo.

La confrontación del modelo con la literatura previa sobre human in the loop permite precisar su aporte. La revisión del campo había identificado que la intervención humana en sistemas educativos inteligentes se concentra en fases de desarrollo y entrenamiento, con presencia marginal en la operación pedagógica cotidiana, y había señalado la imprecisión conceptual de la noción de intervención (Memarian y Doleck, 2024). El modelo HITL-DU responde a esa imprecisión mediante la diferenciación de cinco tipos de intervención cualitativamente distintos, cada uno con su propio indicador de verificación. Esta diferenciación permite distinguir entre configuraciones que satisfacen nominalmente el principio de



supervisión humana y configuraciones que lo satisfacen efectivamente, distinción que la formulación genérica del principio no permite establecer. El indicador propuesto para el tercer momento —proporción de observaciones modificadas o suprimidas por el docente— opera como prueba de existencia de decisión profesional: una tasa de intervención nula no indica un sistema excelente sino, con alta probabilidad, una supervisión que no ocurrió.

Los hallazgos dialogan también con la discusión sobre competencia digital docente. El señalamiento de que este constructo ha sido operacionalizado predominantemente como dominio instrumental encuentra en este trabajo un argumento adicional (Kroff Trujillo y Bellemans Fagalde, 2026). Las competencias que el modelo demanda no son de manejo de herramientas sino de gobierno del juicio evaluativo: capacidad de explicitar criterios, de discriminar relevancia formativa, de detectar error conceptual bajo formulación fluida y de diseñar oportunidades de reelaboración. Se trata de competencias pedagógicas de alto nivel cuya adquisición no se resuelve con capacitación instrumental. Esta conclusión converge con los análisis sobre las competencias requeridas al docente investigador en el nuevo escenario tecnológico y con la evidencia sobre percepciones docentes ambivalentes, que combinan expectativa de alivio de carga con temor a la pérdida de control sobre los criterios (Cribilleros-Shigihara et al., 2025; Valenzuela-Velasco, 2026). El modelo ofrece a esa ambivalencia una respuesta operativa: el control no se preserva mediante la abstención, sino mediante la ocupación deliberada de posiciones específicas del ciclo.

La proyección del modelo sobre la evaluación híbrida requiere una precisión. Los estudios sobre modalidades mixtas coinciden en que la integración didáctica no resulta de la disponibilidad de canales sino de decisiones de diseño sobre la articulación de las evidencias producidas en cada uno (Pattier y Ferreira, 2023; Ruiz Domínguez y Maggio, 2025). El modelo HITL-DU es aplicable a escenarios híbridos precisamente porque su primer momento —encuadre— obliga a definir qué evidencia se recoge en qué canal y bajo qué criterios se valora, decisión que en la práctica habitual queda diferida hasta el momento de la calificación. La hibridación, en este sentido, no agrega un problema nuevo al dispositivo sino que vuelve intolerable la ausencia de encuadre que la presencialidad permitía compensar mediante la observación directa.



Desde la perspectiva de la política institucional, los resultados sugieren tres implicaciones. La primera concierne a la regulación: las normativas universitarias sobre uso de inteligencia artificial han tendido a concentrarse en la conducta estudiantil y han desatendido la regulación del uso docente en procesos evaluativos, vacío que deja sin cobertura el escenario de mayor consecuencia para los derechos del estudiante. La segunda concierne a la formación: el desarrollo profesional docente requiere desplazarse desde la alfabetización instrumental hacia el entrenamiento del juicio en condiciones de asistencia algorítmica, lo que supone dispositivos formativos basados en casos y en análisis de salidas defectuosas antes que en tutoriales de uso. La tercera concierne a la carga docente: el modelo no reduce el tiempo de evaluación de manera proporcional al volumen automatizado, porque transfiere tiempo desde la redacción hacia la curaduría y el diseño; las instituciones que adopten sistemas generativos esperando reducciones lineales de carga obtendrán, previsiblemente, degradación de calidad antes que ahorro.

Las limitaciones del estudio condicionan el alcance de estas afirmaciones. El modelo posee estatuto teórico y su formulación deriva de literatura, no de observación directa de prácticas. Los indicadores propuestos requieren operacionalización métrica y verificación de su sensibilidad antes de ser utilizados con fines de auditoría institucional. La ausencia de doble codificación independiente en el análisis limita las garantías de confiabilidad del procedimiento categorial. Finalmente, la velocidad de evolución de las capacidades de los sistemas generativos implica que las delimitaciones establecidas en la Tabla 2 constituyen un estado de situación y no una frontera estable; la actualización periódica de esa matriz es condición de vigencia del modelo. La agenda derivada incluye la validación del modelo mediante estudios cuasiexperimentales que comparen ciclos evaluativos con y sin curaduría estructurada, la construcción de un instrumento de medición de la calidad de la curaduría docente y la exploración de la percepción estudiantil respecto de la transparencia procedimental en devoluciones asistidas.

CONCLUSIONES

El objetivo de construir un modelo conceptual de evaluación híbrida con retroalimentación formativa bajo el principio *humana in the loop*, que explicitara las funciones delegables a sistemas automatizados y aquellas irreductibles a la intervención del profesional de la educación, fue alcanzado mediante la formulación del modelo HITL-DU. El modelo distribuye el ciclo evaluativo en cinco momentos — encuadre, generación asistida, curaduría docente, devolución dialógica y metaevaluación— y asigna a



cada uno un rol profesional específico, un indicador de calidad verificable y un riesgo asociado a la delegación total.

El análisis del corpus permitió establecer que la retroalimentación constituye un punto de estrangulamiento estructural de la evaluación universitaria, determinado por condiciones materiales de ejercicio antes que por concepciones pedagógicas del profesorado, y que la hibridación de las modalidades de enseñanza intensifica esa restricción al distribuir las evidencias de aprendizaje entre canales de distinta capacidad de captura. Se estableció, asimismo, que los sistemas generativos presentan un perfil de desempeño asimétrico, con ventajas consistentes en cobertura, velocidad y extensión, y limitaciones persistentes en diagnóstico conceptual, calibración de la exigencia y pertinencia contextual, junto con sesgos documentados en la valoración de producción escrita.

La conclusión de mayor alcance concierne a la naturaleza de la contribución docente. La evidencia analizada indica que el valor pedagógico de la herramienta no es una propiedad de la herramienta sino un efecto de la arquitectura de decisión que el docente construye alrededor de ella. Dos instituciones que utilicen el mismo sistema con el mismo grupo de estudiantes obtendrán resultados formativos divergentes según ocupen o no las posiciones de encuadre, curaduría y metaevaluación. Esta constatación reposiciona la competencia digital docente: no se trata de una destreza instrumental adicional al repertorio profesional, sino de una competencia de gobierno del juicio evaluativo que moviliza los saberes pedagógicos de mayor complejidad.

Se derivan de lo anterior recomendaciones en tres planos. En el plano de la práctica docente, se recomienda no incorporar sistemas generativos al ciclo evaluativo sin rúbrica validada previa y sin previsión explícita de una instancia de reelaboración estudiantil con efecto sobre la valoración final. En el plano institucional, se recomienda extender las normativas sobre uso de inteligencia artificial al ámbito de las prácticas evaluativas docentes y adoptar indicadores de verificación de la curaduría antes que declaraciones genéricas de supervisión humana. En el plano de la formación del profesorado, se recomienda reorientar los dispositivos de desarrollo profesional desde la capacitación instrumental hacia el entrenamiento del juicio evaluativo en condiciones de asistencia algorítmica, mediante análisis de casos y examen crítico de salidas defectuosas.



El modelo propuesto requiere validación empírica. Su formulación teórica constituye un punto de partida para la investigación en didáctica universitaria latinoamericana, campo en el que la adopción tecnológica ha precedido sistemáticamente a la formulación de criterios pedagógicos. Invertir esa secuencia —establecer primero qué decisiones son irrenunciablemente humanas y luego determinar qué puede asistirse— constituye, más allá del dispositivo particular aquí propuesto, la orientación estratégica que la evidencia disponible respalda.

REFERENCIAS BIBLIOGRÁFICAS

- Albarrán Martínez, A. A. (2026). Inteligencia artificial generativa y planeación didáctica: implicaciones en la práctica docente. *Revista de Estudios en Práctica Docente e Innovación Educativa*, 1(1), 57-76. <https://doi.org/10.24844/enlt/2604>
- Bennasar-García, M. I., y Estrada, J. M. (2021). La evaluación de los aprendizajes, en la dicotomía educativa presencial-virtual. *Horizontes. Revista de Investigación en Ciencias de la Educación*, 5(18), 434-446. <https://doi.org/10.33996/revistahorizontes.v5i18.186>
- Black, P., y Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Bonnefoy Valdés, N. (2021). Evaluación de competencias en educación superior: conceptos, principios y agentes. *Revista Educación*, 45(1), 578-592. <https://doi.org/10.15517/revedu.v45i1.43444>
- Boud, D., y Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698-712. <https://doi.org/10.1080/02602938.2012.691462>
- Carless, D., y Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315-1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Cordón García, O. (2023). Inteligencia artificial en educación superior: oportunidades y riesgos. *RiiTE Revista Interuniversitaria de Investigación en Tecnología Educativa*, 16-27. <https://doi.org/10.6018/riite.591581>



- Cribilleros-Shigihara, O. A., Gamarra-Chirinos, O. P., Vergara-Vasquez, A. T., y Villanueva-Javes, J. A. (2025). Competencias del docente investigador en la era de la inteligencia artificial: retos y oportunidades en educación superior. *Revista de Ciencias Sociales*, 31(4), 517-528.
<https://doi.org/10.31876/rcs.v31i4.44878>
- Cui, Y. (2026). Artificial intelligence and human assessment consistency and bias in higher education academic writing evaluation. *Discover Artificial Intelligence*, 6(1).
<https://doi.org/10.1007/s44163-026-01286-0>
- Dávila Ramírez, J. R., y Huertas Martínez, J. A. (2024). Evaluación formativa y calidad de la retroalimentación: diseño y validación de escalas para profesores escolares. *Educación XXI*, 27(2), 167-194. <https://doi.org/10.5944/educxx1.38283>
- Del Valle, E. (2025). La inteligencia artificial en la educación: potencial transformador, riesgos de sesgo y desafíos éticos. *Revista Iberoamericana de Educación*, 99(1), 79-93.
<https://doi.org/10.35362/rie9916838>
- Gómez Reyes, A. (2023). Intervenciones de retroalimentación formativa. *PARADIGMA*, 16-35.
<https://doi.org/10.37618/paradigma.1011-2251.2023.p16-35.id1465>
- Hattie, J., y Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Huerta Rosales, M., y Vigoa Escobedo, Y. (2024). La evaluación formativa desde el enfoque de la evaluación auténtica en la educación superior universitaria. *Revista de Investigación Científica Huamachuco*, 2(2), 91-105. <https://doi.org/10.61709/4akj9s23>
- Jacobs, J., Scornavacco, K., Harty, C., Suresh, A., Lai, V., y Sumner, T. (2022). Promoting rich discussions in mathematics classrooms: using personalized, automated feedback to support reflection and instructional change. *Teaching and Teacher Education*, 112, 103631.
<https://doi.org/10.1016/j.tate.2022.103631>
- Jin, F. J.-Y., Maheshi, B., Lai, W., Li, Y., Gasevic, D., Chen, G., Charwat, N., Chan, P. W. K., Martinez-Maldonado, R., Gasevic, D., y Tsai, Y.-S. (2025). Students' perceptions of generative AI-powered learning analytics in the feedback process. *Journal of Learning Analytics*, 12(1), 152-168. <https://doi.org/10.18608/jla.2025.8609>



- Kroff Trujillo, F. J., y Bellemans Fagalde, P. J. (2026). La competencia digital docente en educación superior: tensiones, enfoques y desafíos contemporáneos. *Revista TSup (Transformación Superior)*, 2(1), 68-84. <https://doi.org/10.53382/issn.2810-7977.41>
- Lee, S. S., y Moore, R. L. (2024). Harnessing generative AI (GenAI) for automated feedback in higher education: a systematic review. *Online Learning*, 28(3). <https://doi.org/10.24059/olj.v28i3.4593>
- Leiva-Guerrero, M. V., Araya Zamorano, I., Escobar Collins, R., y Silva Castro, F. (2026). Retroalimentación de aprendizajes con inteligencia artificial generativa en estudiantes universitarios. *RIED-Revista Iberoamericana de Educación a Distancia*, 29(1), 241-267. <https://doi.org/10.5944/ried.45547>
- Memarian, B., y Doleck, T. (2024). Human-in-the-loop in artificial intelligence in education: a review and entity-relationship (ER) analysis. *Computers in Human Behavior: Artificial Humans*, 2(1), 100053. <https://doi.org/10.1016/j.chbah.2024.100053>
- Mendonça, P. C., Quintal, F., Figueiredo, M., Baras, K., y Mendonça, F. (2026). A hybrid reasoning framework for artificial intelligence assessment rubric generation in human and automated contexts: evidence from an undergraduate programming course. *Computers and Education: Artificial Intelligence*, 11, 100663. <https://doi.org/10.1016/j.caeai.2026.100663>
- Muñoz Cortés, P., Henríquez D'Aquino, E., y Kunakov Pérez, N. (2023). Retroalimentación como evaluación formativa desde la perspectiva docente en odontología: estudio de caso. *Educación Médica*, 24(2), 100785. <https://doi.org/10.1016/j.edumed.2022.100785>
- Ninaus, M., y Sailer, M. (2022). Closing the loop - the human role in artificial intelligence for education. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.956798>
- Parra Martínez, B. D., Orejarena Silva, H. A., y Acosta Salazar, H. U. (2022). Retroalimentación docente universitaria. *Inclusión y Desarrollo*, 9(1), 24-36. <https://doi.org/10.26620/uniminuto.inclusion.9.1.2022.24-36>
- Pattier, D., y Ferreira, P. D. (2023). ¿Presencial, online o híbrida?: percepciones y preferencias por parte de docentes y estudiantes de educación superior. *Revista Complutense de Educación*, 34(1), 69-79. <https://doi.org/10.5209/rced.76766>



- Ruiz Domínguez, M. Á., y Maggio, M. (2025). Evaluación de una didáctica en vivo híbrida en educación superior: un estudio de caso. REDU. Revista de Docencia Universitaria, 23(1), 77-92.
<https://doi.org/10.4995/redu.2025.22702>
- Santana-Soriano, E. (2025). Ética e inteligencia artificial en educación: un análisis bibliométrico de tendencias emergentes y desafíos. Ciencia y Educación, 9(1), 95-111.
<https://doi.org/10.22206/cyed.2025.v9i1.3255>
- Tam, A. C. F. (2024). Evolution of feedback ecology: reshaping students' conceptions and practices in feedback literacy. Assessment & Evaluation in Higher Education, 49(3), 376-391.
<https://doi.org/10.1080/02602938.2023.2246701>
- Valdez Valdez, L. S., Sánchez Uscamayta, J. O., y Lescano López, G. S. (2023). Evaluación formativa: retroalimentación, estrategias e instrumentos. Revista Educación, 47(2), 1-24.
<https://doi.org/10.15517/revedu.v47i2.53987>
- Valenzuela-Velasco, C. A. (2026). Percepción docente sobre el uso de la inteligencia artificial en la educación superior. Prohominum, 8(2), 473-482. <https://doi.org/10.47606/acven/ph0503>
- Viberg, O., Poquet, O., Kovanovic, V., y Khosravi, H. (2025). Fostering human agency in age of AI. Journal of Learning Analytics, 12(3), 1-7. <https://doi.org/10.18608/jla.2025.9485>
- Woitt, S., Weidlich, J., Jivet, I., Orhan Göksün, D., Drachsler, H., y Kalz, M. (2025). Students' feedback literacy in higher education: an initial scale validation study. Teaching in Higher Education, 30(1), 257-276. <https://doi.org/10.1080/13562517.2023.2263838>
- Yeung, S. (2025). University students' engagement with generative AI-supported automated writing evaluation (AWE) feedback. Journal of Second Language Writing, 68, 101203.
<https://doi.org/10.1016/j.jslw.2025.101203>
- Zhan, Y., y Yan, Z. (2026). Students' engagement with ChatGPT feedback: implications for student feedback literacy in the context of generative artificial intelligence. Assessment & Evaluation in Higher Education, 51(5), 821-834. <https://doi.org/10.1080/02602938.2025.2471821>

