

Anexo 2 INNOVA
MEMORIA FINAL. COMPROMISOS Y RESULTADOS
PROYECTOS DE INNOVACIÓN Y MEJORA DOCENTE
2025/2026

Identificación del proyecto	
Código	sol-202500299713-tra
Título	Spreadify: sistema de aprendizaje de contabilidad haciendo uso de hojas de cálculo e inteligencia artificial (actualmente ContaMIND)
Responsable	José Miguel Mota Macías

1. Describa los resultados obtenidos a la luz de los objetivos y compromisos que adquirió en la solicitud de su proyecto. Incluya tantas tablas como objetivos contempló.

Objetivo nº 1	<i>Integrar capacidades de Inteligencia Artificial para la personalización adaptativa del aprendizaje.</i>
Actividades que había previsto en la solicitud del proyecto:	• <i>Desarrollar un módulo de IA que analice los patrones de uso y errores del alumnado en Spreadify.</i> • <i>Crear un sistema automatizado para generar y proponer ejercicios de refuerzo personalizados basados en las necesidades detectadas por la IA.</i> • <i>Establecer un banco de ejercicios de refuerzo clasificados que la IA pueda utilizar para la personalización.</i>
Actividades realizadas y resultados obtenidos:	Replanteamiento tecnológico. Para integrar la Inteligencia Artificial (IA) con garantías de privacidad y trazabilidad, Spreadify evolucionó desde su versión anterior, basada en hojas de cálculo de Google Workspace, a una plataforma web propia desplegada en un servidor de la Universidad de Cádiz (https://sibila.uca.es/spreadify/). Durante el curso, se ha registrado la herramienta con el nombre ContaMIND, ya que la denominación original coincidía con la de otro producto existente. La nueva arquitectura conserva la práctica con retroalimentación inmediata del proyecto anterior y añade las capas de IA y analítica que se describen a continuación. Módulo de análisis de patrones de uso y errores. La plataforma registra cada acción del alumnado (altas, modificaciones y eliminaciones de líneas, solicitudes de revisión y consultas a la IA), y un motor de revisión automática compara cada asiento con la solución definida por el profesorado. El motor distingue los asientos descuadrados de los que cuadran pero son incorrectos, y clasifica los errores en tres tipos: de cuenta (48 %), de importe (33 %) y estructurales (19 %). En el periodo analizado se registraron 103.563 acciones y 390.828 cambios a nivel de campo. Tutor de IA con generación aumentada por recuperación (RAG) controlada. Se desarrolló un asistente, denominado Pacioli por Luca Pacioli, padre de la contabilidad moderna, que responde exclusivamente a partir de los materiales que sube el profesorado. Funciona con modelos de lenguaje ejecutados localmente en el servidor de la UCA, con generación determinista, más de veinte restricciones explícitas (no inventar cuentas ni importes, no recurrir a conocimiento externo), un mensaje de rechazo fijo cuando el material no contiene la respuesta y el registro de todas las interacciones para su auditoría (Figura 1). Se implementó además un mecanismo pedagógico de control: cuando el

alumnado alcanza el número de revisiones fijado por el profesor, el botón de revisión se bloquea hasta que consulta al tutor. Así, la IA actúa como andamiaje y se evita la corrección por ensayo y error.

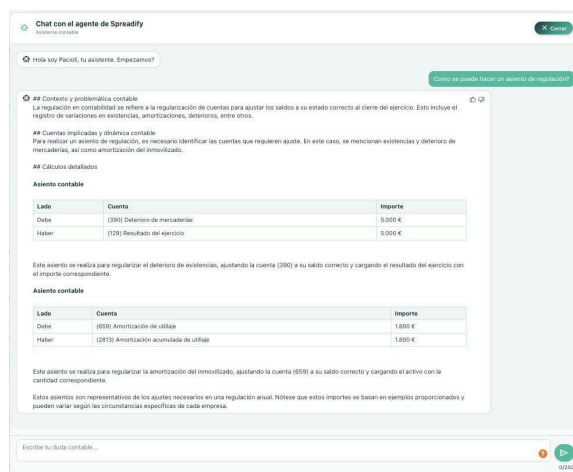


Figura 1. Tutor de IA respondiendo a partir de los materiales del curso.

Banco de ejercicios. Se creó un repositorio con etiquetas y filtros que admite ciclos contables completos o ejercicios que parten de un balance de sumas y saldos intermedio, y que permite compartir ejercicios entre el profesorado. En el curso 2025/2026 se programaron 12 ejercicios en los dos grupos participantes.

Desajuste y solución adoptada. No se implementó la generación automática de ejercicios personalizados mediante IA. Durante el desarrollo se constató que los modelos generativos pueden producir asientos contables erróneos, un riesgo que señala la literatura sobre IA en educación contable (Ballantine et al., 2024; Wiseman y Foster, 2025). Por ello se priorizó un tutor controlado y auditable, y la personalización se obtiene mediante la retroalimentación por asiento y la contextualización del tutor en cada ejercicio. La generación de ejercicios con validación docente previa queda como línea de continuidad.

Objetivo nº 2	<i>Integrar Spreadify con el sistema de evaluación académica de la asignatura.</i>
Actividades que había previsto en la solicitud del proyecto:	● <i>Diseñar un mecanismo de sincronización que permita la transferencia automática de las calificaciones de los ejercicios resueltos en Spreadify al sistema de evaluación universitario.</i> ● <i>Configurar criterios de ponderación para que la resolución de ejercicios seleccionados en Spreadify contribuya directamente a la nota final de la asignatura.</i> ● <i>Implementar pruebas de seguridad y exactitud para asegurar la fiabilidad de la integración de datos.</i>
Actividades realizadas y resultados obtenidos:	Sincronización con el sistema de evaluación. Durante el desarrollo se replanteó el alcance de este objetivo. La transferencia automática de las calificaciones generadas en la plataforma planteó reparos del comité ético, ya que la intervención de sistemas de IA en la calificación final supone una decisión automatizada con efectos académicos (art. 22 del Reglamento General de Protección de Datos). Además, el Reglamento Europeo de Inteligencia Artificial considera de alto riesgo los sistemas destinados a evaluar resultados de aprendizaje. Por ello, ContaMIND se mantiene como herramienta de aprendizaje formativo, sin intervención en la evaluación sumativa, y la calificación es responsabilidad exclusiva del profesorado.

	Criterios de ponderación. La participación en las sesiones prácticas con ContaMIND supuso un 20 % de la calificación final de la asignatura. Se acreditó mediante hoja de firmas en el aula, de modo que la calificación queda desvinculada tanto de la puntuación obtenida en la herramienta como de los registros de actividad de la plataforma. Así se incentiva el uso continuado de ContaMIND sin que sus resultados automáticos condicionen la evaluación. Pruebas de seguridad y exactitud. Esta actividad se completó íntegramente. El algoritmo de revisión automática se validó con una batería de 27 casos de prueba sobre nueve condiciones de comportamiento (coincidencia exacta, tolerancia en importes, orden de las líneas, agregación de líneas, cuentas erróneas y líneas omitidas o sobrantes, entre otras), todos superados sin falsos positivos ni negativos. La plataforma funciona en un servidor de la UCA con autenticación por roles, comunicaciones cifradas y modelos de IA locales, sin cesión de datos a servicios externos. El tratamiento de datos se comunicó al Delegado de Protección de Datos, se incorporó un aviso informativo obligatorio en el primer acceso y los análisis se realizaron sobre datos anonimizados. Valoración. El replanteamiento refuerza el carácter formativo de la herramienta y garantiza un uso de la IA en la docencia ético y conforme a la normativa.
--	--

Objetivo nº 3	<i>Mejorar la funcionalidad del cuadro de mando (dashboard) para ofrecer análisis de rendimiento más profundos y comparativos.</i>
Actividades que había previsto en la solicitud del proyecto:	● <i>Ampliar las métricas y visualizaciones del dashboard para incluir un análisis más detallado del progreso individual y grupal.</i> ● <i>Incorporar filtros y opciones de personalización que permitan a profesores y alumnos explorar los datos de rendimiento de manera más granular.</i> ● <i>Integrar los resultados del análisis de la LA en el dashboard para mostrar los puntos débiles detectados y las sugerencias de refuerzo personalizadas.</i>
Actividades realizadas y resultados obtenidos:	Dashboard del profesorado. Se amplió el panel docente con indicadores de grupo (alumnado activo, ejercicios programados, porcentaje que inicia y que completa, nota media y mediana, comparativa con el curso), distribución de notas, actividad diaria, calendario de ejercicios y un apartado de «hallazgos clave» generado automáticamente, que señala, por ejemplo, el ejercicio con la media más baja o el número de estudiantes con señales de bajo progreso (Figura 2). Incluye además el detalle por estudiante, la comparación entre su respuesta y la solución, y un mapa de calor por estudiante y ejercicio.

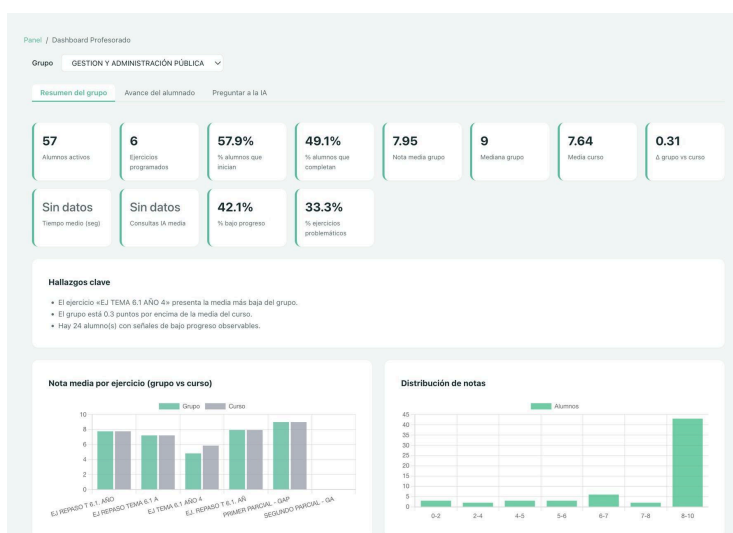


Figura 2. Dashboard del profesorado (grupo de Gestión y Administración Pública).

Dashboard del alumnado. Cada estudiante dispone de un panel con sus ejercicios programados, iniciados y finalizados, su nota en cada ejercicio frente a la media del grupo (percentil y diferencia), el tiempo dedicado, el número de consultas a la IA, la evolución de sus puntuaciones y un diagnóstico de fortalezas y puntos de mejora (Figura 3).



Figura 3. Dashboard de progreso del alumnado.

Filtros y personalización. Todos los indicadores pueden filtrarse por grupo, ejercicio y rango de fechas. En febrero de 2026, funcionalidades como la calificación y el progreso del estudiante o la consulta de sus soluciones figuraban como pendientes; al cierre del cuatrimestre estaban operativas.

Integración de los resultados de la IA. El dashboard recoge el uso del tutor (consultas por estudiante y ejercicio) y los diagnósticos de error del motor de revisión. Se implementó también un módulo de alertas configurable con diez métricas (puntuación mínima, revisiones sin avance, exceso de consultas a la IA, asientos descuadrados o abandono de ejercicios, entre otras), que se mantuvo desactivado de forma deliberada en este primer despliegue para calibrar los umbrales con la práctica docente y evitar la saturación de avisos. Las sugerencias de refuerzo personalizadas quedan pendientes para el próximo curso, vinculadas

a la activación de este módulo.

- Realice una breve valoración sobre la influencia del proyecto ejecutado en la evolución de las asignaturas implicadas.

Análisis del impacto de la innovación en las asignaturas relacionadas con el proyecto

ContaMIND se implantó en el segundo cuatrimestre del curso 2025/2026 en **Fundamentos de Contabilidad Financiera**, asignatura obligatoria de primer curso que se imparte en la Facultad de Ciencias Sociales y de la Comunicación (campus de Jerez) en dos titulaciones: el Grado en Gestión y Administración Pública (57 matriculados) y el Grado en Marketing y Publicidad (140). La actividad con la plataforma se desarrolló entre el 6 de abril y el 21 de mayo de 2026 en sesiones prácticas presenciales.

Evolución respecto al proyecto anterior. La Tabla 1 compara esta edición con la de 2024/2025.

Tabla 1. Comparativa entre las dos últimas ediciones del proyecto.

Aspecto	Edición 2024/2025	Edición 2025/2026
Tecnología	Hojas de cálculo en Google Workspace	Plataforma web propia en un servidor de la UCA (ContaMIND)
Retroalimentación	Por celda: valor y fórmula (verde, amarillo, rojo)	Por asiento, con diagnóstico de cuadre y tipo de error
Inteligencia artificial	No incluida	Tutor RAG con modelos locales y control del número de revisiones
Análítica docente	Paneles en hojas de cálculo	Dashboards de profesorado y alumnado, con filtros y alertas configurables
Volumen de uso registrado	1.603 registros de uso y 1.585 envíos	103.563 acciones y 9.735 revisiones automáticas

Uso e implicación del alumnado. De los 194 estudiantes, 118 (60,8 %) realizaron al menos un intento: en total, 464 intentos sobre 12 ejercicios, 30.949 líneas contables introducidas y 9.735 revisiones automáticas (22,6 por intento de media). Las 133.497 ediciones registradas frente a esas 30.949 líneas reflejan un trabajo iterativo de corrección, no de entrega única.

Progreso dentro de la plataforma. La puntuación media pasó del 11,3 % en la primera revisión al 69,5 % en la última (la mediana, del 4,5 % al 81,8 %). Mejoraron 354 de los 380 intentos revisados y ninguno empeoró (prueba de Wilcoxon, $p < 0,001$) (Figura 4). El número de revisiones y de ediciones se asoció positivamente con la puntuación final (ρ de Spearman = 0,62 y 0,72, respectivamente). Este patrón es coherente con la literatura sobre retroalimentación formativa y experiencias de dominio (Shute, 2008; Van Dinther et al., 2011), aunque se interpreta como indicador de proceso y no como efecto causal sobre el aprendizaje.

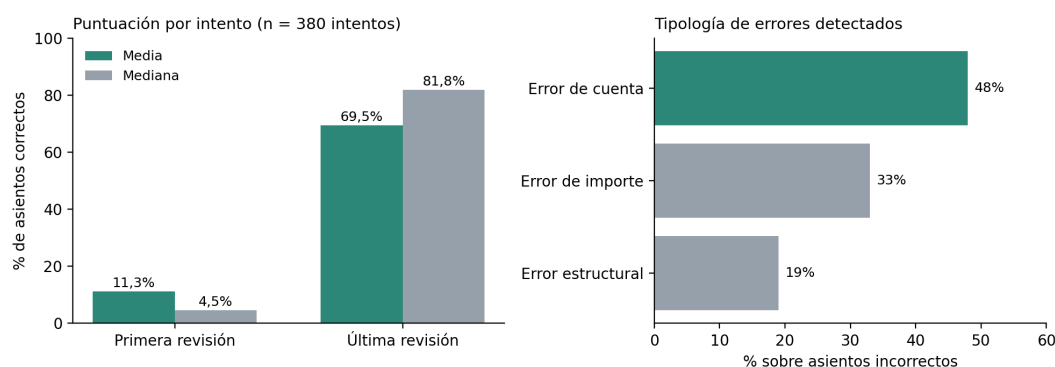


Figura 4. Puntuación en la primera y la última revisión, y tipología de los errores detectados.

Dificultades detectadas y ajustes docentes. El 98 % de las observaciones de asientos estaban

cuadradas, pero solo el 46,7 % eran correctas: 121.461 asientos cuadraban y aun así eran incorrectos. Es decir, el alumnado domina la mecánica de la partida doble, pero falla al elegir la cuenta, un patrón coherente con el concepto umbral de dualidad descrito en la educación contable (Lucas y Mladenovic, 2006). Los errores de cuenta se concentraron en los ejercicios con descuentos, y la puntuación final media osciló entre el 53,6 % y el 89,7 % según el ejercicio. Esta información, visible en el dashboard, permite al profesorado reforzar en clase la selección de cuentas y el tratamiento de descuentos e IVA, y ofrece un criterio objetivo para revisar la secuencia de ejercicios.

Cambio metodológico. La asignatura incorpora de forma estable la práctica autónoma con corrección inmediata, un tutor que orienta sin dar la solución y un seguimiento docente basado en datos, tres mecanismos alineados con la evidencia sobre aprendizaje activo en contabilidad (Freeman et al., 2014; Gittings et al., 2020). La herramienta se ha utilizado como complemento de la docencia presencial y no como sustituto de esta.

3. Incluya en la siguiente tabla el número de alumnado matriculado y el de respuestas recibidas en cada opción y realice una valoración crítica sobre la influencia que el proyecto ha ejercido en la opinión del alumnado.

Opinión del alumnado al inicio del proyecto				
Número de alumnado matriculado:		197 alumnos/as (57 GAP y 140 Marketing y Publicidad)		
Valoración del grado de dificultad que cree que va a tener en la comprensión de los contenidos y/o en la adquisición de competencias asociadas a la asignatura en la que se enmarca el proyecto de innovación docente				
Ninguna dificultad	Poca dificultad	Dificultad media	Bastante dificultad	Mucha dificultad
n.d.	n.d.	n.d.	n.d.	n.d.
Opinión del alumnado en la etapa final del proyecto				
Valoración del grado de dificultad que ha tenido en la comprensión de los contenidos y/o en la adquisición de competencias asociadas a la asignatura en la que se enmarca el proyecto de innovación docente				
Ninguna dificultad	Poca dificultad	Dificultad media	Bastante dificultad	Mucha dificultad
n.d.	n.d.	n.d.	n.d.	n.d.
Los elementos de innovación y mejora docente aplicados en esta asignatura han favorecido mi comprensión de los contenidos y/o la adquisición de competencias asociadas a la asignatura				
Nada de acuerdo	Poco de acuerdo	Ni en acuerdo ni en desacuerdo	Muy de acuerdo	Completamente de acuerdo
n.d.	n.d.	n.d.	n.d.	n.d.
En el caso de la participación de profesorado invitado				
La participación del profesorado invitado ha supuesto un gran beneficio en mi formación				
Nada de acuerdo	Poco de acuerdo	Ni en acuerdo ni en desacuerdo	Muy de acuerdo	Completamente de acuerdo
No procede	No procede	No procede	No procede	No procede
Valoración crítica sobre la influencia que ha ejercido el proyecto en la opinión del alumnado				
Nota: n.d. = dato no disponible. La tabla solo recoge el número de alumnado matriculado porque en esta				

edición no se aplicó el cuestionario de percepción inicial y final que propone la plantilla. La opinión del alumnado se ha analizado a partir de tres fuentes: un diagnóstico previo, los indicadores de uso de la plataforma y los resultados de la edición anterior.

Diagnóstico previo. Antes de la implantación de ContaMIND se aplicó un cuestionario a 94 estudiantes de 1.º y 2.º curso de contabilidad (21 ítems Likert de 1 a 5, agrupados en cuatro dimensiones), cuyos resultados se presentaron en el XXII Encuentro Internacional ASEPUC (Anexo I). El alumnado valora positivamente el aprendizaje de la contabilidad ($M = 3,86$) y muestra una fuerte preferencia por las metodologías prácticas ($M = 4,45$): el 92,6 % prefiere actividades prácticas en clase y el 83 % considera que las simulaciones ayudan a entender la materia. La competencia percibida para interpretar y utilizar información financiera es la dimensión más débil ($M = 3,62$) (Figura 5). En las respuestas abiertas, el 60,6 % pide más ejercicios y casos prácticos, y el 18,2 %, más tecnología o IA. Este diagnóstico orientó el diseño de ContaMIND hacia la práctica guiada, la retroalimentación inmediata y el seguimiento docente.

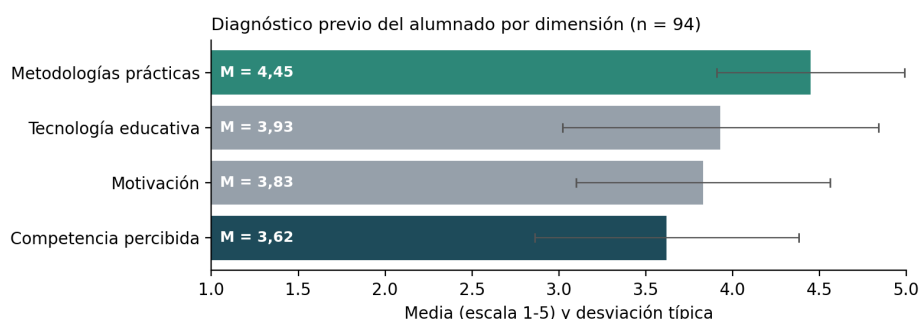


Figura 5. Diagnóstico previo del alumnado por dimensión (n = 94).

Respuesta del alumnado a la herramienta. La implicación se ha medido con indicadores objetivos: el 60,8 % del alumnado utilizó la plataforma, con una media de 22,6 revisiones por intento, y la puntuación media subió del 11,3 % al 69,5 % entre la primera y la última revisión. Las consultas al tutor se centraron sobre todo en el tratamiento de descuentos e IVA, y ninguna pidió resolver el ejercicio completo, lo que apunta a un uso del tutor como apoyo y no como atajo. Estos datos deben leerse con cautela: la participación puntuaba en la calificación, el diseño no incluye grupo de control y el tutor tuvo un uso limitado (153 consultas de 30 estudiantes), lo que indica que conviene integrarlo mejor en las sesiones.

Antecedente. En la edición 2024/2025 del proyecto, 31 de los 33 estudiantes encuestados (94 %) manifestaron estar muy o completamente de acuerdo en que la herramienta favoreció su comprensión de los contenidos.

Limitación y mejora. La falta de una medición final de la percepción es la principal limitación de esta edición. Para resolverla, el cuestionario de la plantilla se integrará en la propia plataforma y se mostrará al alumnado en su primer y en su último acceso.

4. Describa las medidas de difusión a las que se comprometió en la solicitud y las que ha llevado a cabo. Describe, analice y valore los desajustes respecto a lo previsto en el proyecto y las soluciones aplicadas.

Descripción de las medidas comprometidas en la solicitud
<i>Seminarios para profesores: se realizarán seminarios y talleres para el profesorado de la universidad, compartiendo las experiencias y resultados del proyecto. Estos seminarios facilitarán la adopción de tecnologías similares en otras asignaturas y departamentos.</i>
Descripción de las medidas que se han llevado a cabo
Las acciones realizadas, en orden cronológico, han sido las siguientes: • Demostración al Departamento de Economía Financiera y Contabilidad (13 de febrero de 2026). Presentación de la herramienta al profesorado del departamento. Como resultado, se

incorporaron al proyecto los profesores Jesús Herrera Madueño y Paula Isabel Rodríguez Castro.

- **Instituto de Censores Jurados de Cuentas de España y Consejo General de Economistas de España (marzo de 2026).** Presentación a cargo del profesor Jesús Herrera Madueño, que traslada la experiencia al ámbito profesional contable y auditor.
- **Registro del software** por la Universidad de Cádiz en Safe Creative con el nombre ContaMIND (identificador 2605085574200, 8 de mayo de 2026) (Anexo III).
- **XXII Encuentro Internacional ASEPUC (Valencia, 10-12 de junio de 2026).** Comunicación «Herramientas tecnológicas en la formación contable: el caso de Spreadify» (Anexo I).
- **Artículo en proceso de revisión** en la revista *Education and Information Technologies* (Springer, Q1): «ContaMIND: an integrated IT architecture for automated formative feedback, learning analytics, and controlled RAG tutoring in accounting education» (Anexo II).
- **Seminario de presentación de resultados (25 de septiembre de 2026),** en el seminario 1.750 de la Facultad de Ciencias Sociales y de la Comunicación, abierto a todo el profesorado de la facultad.
- **Plataforma accesible** en <https://sibila.uca.es/spreadify/> para el profesorado interesado en utilizarla.

Análisis de los desajustes y soluciones aplicadas

El compromiso de la solicitud se ha cumplido y ampliado. Se previeron seminarios para el profesorado de la UCA, y la difusión ha alcanzado tres niveles: institucional (departamento y facultad), profesional (Instituto de Censores Jurados de Cuentas y Consejo General de Economistas) e internacional (encuentro ASEPUC y revista Q1).

Se han producido dos desajustes respecto a lo previsto, ninguno con efecto negativo sobre los resultados. El primero es de calendario: el seminario abierto a toda la facultad se celebró al final del periodo, en septiembre de 2026, para presentar los resultados completos del despliegue. Para compensarlo, en febrero se realizó una demostración al departamento que, además de difundir la herramienta, amplió el equipo del proyecto.

El segundo afecta a la financiación. El proyecto recibió una ayuda de 300 € para la presentación o publicación de los resultados, que finalmente no fue necesario utilizar: la participación en el XXII Encuentro Internacional ASEPUC no generó gastos con cargo al proyecto, y los resultados se han orientado a su publicación en una revista de impacto (*Education and Information Technologies*, Springer, Q1), cuyo proceso de revisión sigue en curso. En consecuencia, la ayuda no se ha ejecutado. La publicación del artículo completará la difusión de los resultados del proyecto.

5. **Conclusiones.** Entre ellas, describa las aplicaciones futuras del proyecto, su sostenibilidad y posible continuidad y estrategias para ello.

Conclusiones

Balance. El proyecto ha transformado Spreadify en ContaMIND, una plataforma propia de la UCA que integra simulación contable, revisión automática por asiento, un tutor de IA controlado y analítica para profesorado y alumnado. Los objetivos 1 y 3 se han cumplido en lo esencial, con dos funcionalidades reorientadas o aplazadas de forma justificada (la generación automática de ejercicios y las sugerencias de refuerzo en el dashboard). El objetivo 2 se ha replanteado por criterios éticos y normativos, manteniendo la ponderación por participación y completando las pruebas de seguridad y exactitud.

Sostenibilidad. La continuidad del proyecto se apoya en cuatro factores. La plataforma está alojada en infraestructura de la UCA y no depende de servicios comerciales externos. El software está registrado a nombre de la universidad. El equipo se ha ampliado a cinco profesores de dos departamentos (Economía Financiera y Contabilidad e Ingeniería Informática), con un perfil interdisciplinar. Y el proyecto acumula tres ediciones consecutivas (2023/2024, 2024/2025 y 2025/2026), cada una construida sobre los resultados de la anterior.

Aplicaciones futuras y estrategias de continuidad. Para el curso 2026/2027 se plantean las siguientes líneas:

- Integrar en la plataforma el cuestionario de percepción inicial y final, que se mostrará en el primer y el último acceso para garantizar una medición completa de la opinión del alumnado.
- Activar el módulo de alertas con umbrales calibrados y vincularlo a sugerencias de refuerzo en el dashboard.
- Incorporar clases de equivalencia de cuentas, para aceptar alternativas válidas del Plan General Contable y reducir los falsos errores.
- Evaluar la precisión y la calidad pedagógica del tutor de IA mediante un banco de preguntas validado por expertos.
- Aplicar un diseño cuasi experimental con grupo de comparación y medidas previas y posteriores, que permita contrastar el efecto sobre el rendimiento académico.
- Extender ContaMIND a otras asignaturas y titulaciones de la UCA, aprovechando el interés generado en el departamento y en la facultad, y explorar su transferencia a otros ámbitos procedimentales, como la auditoría o la fiscalidad, junto con las instituciones profesionales a las que se ha presentado.
- Estudiar la generación de ejercicios de refuerzo asistida por IA, siempre con validación docente previa.

Referencias bibliográficas

- Ballantine, J., Boyce, G. y Stoner, G. (2024). A critical review of AI in accounting education: Threat and opportunity. *Critical Perspectives on Accounting*, 99, 102711. <https://doi.org/10.1016/j.cpa.2024.102711>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H. y Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415. <https://doi.org/10.1073/pnas.1319030111>
- Gittings, L., Taplin, R. y Kerr, R. (2020). Experiential learning activities in university accounting education: A systematic literature review. *Journal of Accounting Education*, 52, 100680. <https://doi.org/10.1016/j.jaccedu.2020.100680>
- Lucas, U. y Mladenovic, R. (2006). Developing new «world views»: Threshold concepts in introductory accounting. En J. H. F. Meyer y R. Land (Eds.), *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge* (pp. 148-159). Routledge.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189. <https://doi.org/10.3102/0034654307313795>
- Van Dinther, M., Dochy, F. y Segers, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review*, 6(2), 95-108. <https://doi.org/10.1016/j.edurev.2010.10.003>
- Wiseman, D. y Foster, D. (2025). Bridging AI with cost accounting education: A pedagogical framework using reperformance and critical error review. *Journal of Accounting Education*, 72, 100988. <https://doi.org/10.1016/j.jaccedu.2025.100988>

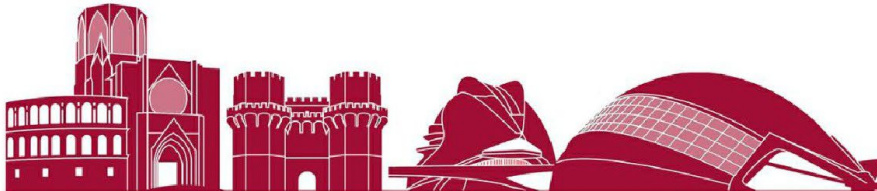
Relación de anexos

Anexo I. Certificado de presentación en el XXII Encuentro Internacional ASEPUC, texto de la comunicación y presentación utilizada.

Anexo II. Artículo remitido a *Education and Information Technologies* (portada y manuscrito).

Anexo III. Certificado de registro del software ContaMIND (Safe Creative).

VALENCIA 2026



XXII ENCUENTRO INTERNACIONAL ASEPUC

El Comité Organizador del **XXII Encuentro Internacional ASEPUC**

CERTIFICA que la comunicación titulada:

HERRAMIENTAS TECNOLÓGICAS EN LA FORMACIÓN CONTABLE: EL CASO DE SPREADIFY

Cuyos autores son:

José Miguel Mota Macías, Jesús Rodríguez Torrejón, María del Rosario Díaz Ortega, Paula Isabel Rodríguez Castro, Jesús Herrera Madueño

Fue presentada en el **XXII Encuentro Internacional ASEPUC**,
celebrado en Valencia del 10 al 12 de junio de 2026



Juan Luis Gandía Cabedo
Co-Presidente del XXII Encuentro
Internacional de ASEPUC



Araceli Mora Enguádanos
Co-Presidenta del XXII Encuentro
Internacional de ASEPUC

Justificación pedagógica de Contamind para la mejora del aprendizaje de la contabilidad en educación superior

Resumen

Este trabajo analiza, desde una perspectiva académica, la pertinencia pedagógica de la herramienta Contamind para mejorar el aprendizaje de la contabilidad en educación superior. El estudio combina dos insumos: (a) un análisis descriptivo previo sobre motivación, metodologías prácticas, competencia percibida y tecnología educativa en una muestra de 94 estudiantes universitarios matriculados en 1.º y 2.º curso; y (b) un documento técnico que define a Contamind como una herramienta educativa apoyada en inteligencia artificial, simulación profesional y analítica docente. Metodológicamente, se adopta un diseño documental-analítico con integración de evidencia empírica local y literatura científica sobre aprendizaje activo, educación contable, autoeficacia académica, feedback y uso pedagógico de la inteligencia artificial. Los resultados muestran una valoración global positiva del aprendizaje de la contabilidad ($M = 3,86/5$), con especial preferencia por las metodologías prácticas ($M = 4,45$) y una menor percepción de competencia para interpretar, diagnosticar y utilizar información financiera en la toma de decisiones ($M = 3,62$). Además, las respuestas abiertas demandan más ejercicios, casos reales, materiales de apoyo y uso significativo de tecnología. A la luz de estos hallazgos, se sostiene que Contamind se encuentra pedagógicamente justificada porque articula tres mecanismos congruentes con las necesidades detectadas: práctica aplicada, retroalimentación inmediata y acompañamiento docente basado en datos. No obstante, el trabajo concluye que su implementación debe ser responsable, supervisada y orientada al desarrollo de juicio profesional, evitando la sustitución acrítica del razonamiento contable.

Palabras clave: educación contable; aprendizaje activo; autoeficacia; inteligencia artificial; feedback; simulación; educación superior.

Abstract

This paper examines the pedagogical relevance of Contamind as a tool to improve accounting learning in higher education. The study integrates two inputs: (a) a prior descriptive analysis of motivation, practical methodologies, perceived competence, and educational technology in a sample of 94 first- and second-year university students; and (b) a technical document defining Contamind as an educational tool supported by artificial intelligence, professional simulation, and teacher analytics. Methodologically, the paper adopts a documentary-analytical design that combines local empirical evidence with scientific literature on active learning, accounting education, academic self-efficacy, feedback, and the pedagogical use of artificial intelligence. Findings show an overall positive evaluation of accounting learning ($M = 3.86/5$), with especially strong support for practical methodologies ($M = 4.45$) and lower perceived competence in interpreting, diagnosing, and using financial information for decision making ($M = 3.62$). Open-ended responses also call for more exercises, real cases, support materials, and meaningful uses of technology. In light of these findings, the paper argues that Contamind is pedagogically justified because it brings together three mechanisms aligned with detected needs: applied practice, immediate feedback, and data-informed teacher support. However, implementation should remain responsible, supervised, and focused on professional judgement rather than replacing accounting reasoning.

Keywords: accounting education; active learning; self-efficacy; artificial intelligence; feedback; simulation; higher education.

1. Introducción

La enseñanza universitaria de la contabilidad afronta un doble desafío. Por un lado, debe garantizar la comprensión de conceptos técnicos, estados financieros y procesos de registro. Por otro, necesita promover transferencia, juicio profesional y capacidad de decisión en contextos aplicados. Esta exigencia se intensifica en un entorno en el que la digitalización, la automatización y la inteligencia artificial están transformando tanto la práctica profesional como las expectativas sobre la formación inicial del alumnado.

En este escenario, la innovación docente no puede justificarse únicamente por novedad tecnológica. Una herramienta educativa solo adquiere legitimidad académica cuando responde a necesidades de aprendizaje bien identificadas y cuando sus funcionalidades son coherentes con mecanismos pedagógicos respaldados por la evidencia. Precisamente desde esta lógica se aborda Contamind: no como un fin en sí mismo, sino como una propuesta didáctica que debe ser analizada a partir de los problemas formativos que pretende resolver.

El presente trabajo presenta dos estructuras bases. La primera trata sobre un análisis descriptivo sobre motivación y aprendizaje de contabilidad en estudiantes de 1.º y 2.º curso, que muestra una preferencia por la práctica, los casos reales y el apoyo tecnológico, pero también revela limitaciones en la competencia percibida para analizar e interpretar información financiera. La segunda parte versa sobre el software Contamind, que se conceptualiza como una herramienta compuesta por simulación profesional, inteligencia artificial adaptativa y panel de seguimiento docente.

Sobre la base anterior, el objetivo del artículo es construir una justificación pedagógica rigurosa del uso de Contamind en educación superior. Se trata, por tanto, de un trabajo de investigación de naturaleza aplicada: no evalúa empíricamente el efecto de la herramienta una vez implantada, sino que examina si su diseño responde de manera razonable y fundamentada a las necesidades detectadas en el alumnado y a lo que la literatura científica considera eficaz en el aprendizaje de la contabilidad.

Tomando como referencia los postulados anteriores, el trabajo pretende analizar las necesidades de aprendizaje detectadas en estudiantes universitarios de la asignatura de contabilidad. Una vez detectada dichas relaciones, el siguiente paso, tras relacionarlas con la literatura previa, será justificar, en términos pedagógicos, el uso de herramientas de apoyo al aprendizaje en Contabilidad.

2. Fundamentación teórica

La evidencia acumulada en educación superior muestra que el aprendizaje activo supera, en promedio, a los formatos exclusivamente expositivos cuando se pretende mejorar el rendimiento académico y reducir el fracaso. En un metaanálisis ampliamente citado, Freeman et al. (2014) documentaron que las metodologías activas producen mejores resultados en exámenes y menores tasas de suspenso que la docencia basada solo en clases magistrales.

En contabilidad, este hallazgo adquiere una relevancia especial. Se trata de una disciplina que no solo exige recordar definiciones o procedimientos, sino interpretar situaciones, detectar errores, formular juicios y explicar implicaciones económicas. En consecuencia, los enfoques basados en ejercicios, casos, simulaciones y tareas aplicadas ofrecen un entorno más adecuado para construir comprensión profunda y transferencia.

La revisión sistemática de Gittings, Taplin y Kerr (2020) sintetiza la investigación sobre actividades de aprendizaje experiencial en educación contable universitaria y concluye que la evidencia disponible es predominantemente positiva, especialmente en términos de conocimiento técnico, comprensión, satisfacción estudiantil y aplicación auténtica de la teoría. Esto sugiere que la formación contable se fortalece cuando el alumnado actúa sobre problemas, datos y decisiones cercanos al ejercicio profesional.

De forma similar, Tran y Herzig (2023) muestran que el aprendizaje basado en casos puede mejorar no solo conocimientos, sino también habilidades profesionales como pensamiento crítico, trabajo en equipo y comunicación. Por ello, cuando los estudiantes demandan más práctica y más casos reales, no están formulando una preferencia superficial, sino apuntando hacia una condición pedagógica importante para aprender mejor.

Otra línea teórica central para este trabajo es la autoeficacia académica. Honicke y Broadbent (2016), en una revisión sistemática de 59 estudios, concluyen que existe una relación positiva moderada entre autoeficacia académica y rendimiento. Asimismo, Van Dinther et al. (2011) sostienen que los programas educativos pueden fortalecer la autoeficacia cuando incorporan experiencias de dominio, feedback y apoyos instruccionales consistentes.

Aplicado a la contabilidad, este marco permite distinguir entre dos planos: reconocer que la asignatura es útil y sentirse realmente capaz de resolver problemas financieros, interpretar balances o apoyar decisiones con información contable. Cuando la autoeficacia es baja o inestable, el alumnado tiende a depender más de la memorización, evita tareas exigentes o muestra menor disfrute en actividades complejas. De ahí la importancia del andamiaje, la práctica progresiva y la retroalimentación inmediata.

Por otro lado, hay que tener presente que la investigación en educación contable ha mostrado que la tecnología añade valor cuando amplía el tiempo de práctica, mejora el acceso a materiales, favorece el repaso y apoya la comprensión de tareas complejas. D'Aquila et al. (2019) encontraron que los vídeos generados por el profesorado mejoraban el rendimiento en una asignatura de principios de contabilidad y eran percibidos como útiles por el alumnado, aunque no como sustitutos de la clase.

En la revisión de la literatura realizada por Theuri et al (2024) sobre una base 187 artículos sobre tecnología y educación contable, observaron que la integración tecnológica constituye un eje consolidado de innovación curricular. Más recientemente, la literatura sobre IA en contabilidad educativa ha subrayado tanto oportunidades como riesgos. Ballantine et al. (2024) argumentan que la IA puede ser una oportunidad para renovar la educación contable, siempre que no reproduzca un enfoque puramente instrumental. Haugland et al (2024) muestran que los estudiantes perciben utilidad en ChatGPT, pero reconocen problemas de fiabilidad y diferencias de uso según el tipo de curso. En una línea más directamente pedagógica, Zhou y Luo (2025) informan que muchos estudiantes perciben que la IA les ayuda con conceptos difíciles y mejora su experiencia de aprendizaje, mientras que Wiseman y Foster (2025) defienden diseños en los que la IA se combina con recálculo manual y revisión crítica de errores para fortalecer juicio profesional y pensamiento analítico.

4. Metodología

El trabajo adopta un diseño documental-analítico con orientación aplicada. No se trata de un experimento ni de una evaluación de impacto de Contamind en condiciones de uso real, sino de una investigación de fundamentación que integra evidencia empírica y literatura científica para analizar la fundamentación pedagógica de la herramienta.

La base empírica procede de un cuestionario previo aplicado a estudiantes universitarios matriculados en 1.º y 2.º curso (véase tabla 1). La muestra final analizada está compuesta por 94 estudiantes. El instrumento incluía 21 ítems Likert (1-5) organizados en cuatro dimensiones - motivación, metodologías prácticas, competencia percibida y tecnología educativa- y una pregunta abierta sobre mejoras deseables en el aprendizaje de la contabilidad.

La segunda fuente empleada es el documento técnico de Contamind, donde la herramienta se describe como un entorno de aprendizaje compuesto por tres módulos: motor de IA inteligente, simulación profesional y panel de seguimiento docente. A partir de esta descripción se realizó un análisis funcional para identificar qué problemas formativos podría abordar cada componente.

En cuanto al tratamiento de la información, se utilizaron los estadísticos descriptivos ya disponibles en el informe previo (medias, desviaciones típicas, distribución por dimensiones y codificación temática de la pregunta abierta) y se contrastaron con hallazgos de la literatura. Este procedimiento permite construir una justificación argumentada, aunque no sustituye a una validación experimental posterior.

Tabla 1. Perfil sintético de la base empírica utilizada

Indicador	Valor
Casos válidos analizados	94
Respuestas originales	97
Casos excluidos (4.º curso)	3
Respuestas abiertas válidas	33
1.º curso	77 (81,9%)
2.º curso	17 (18,1%)
Mujeres	59 (62,8%)
Hombres	34 (36,2%)

5. Resultados

Los resultados empíricos ofrecen un diagnóstico suficientemente claro para orientar una propuesta de innovación docente (Tabla 2). En conjunto, el alumnado valora positivamente el aprendizaje de la contabilidad, pero lo hace con un patrón desigual: la práctica aplicada aparece como la principal fortaleza, mientras que la confianza para interpretar y utilizar información financiera de manera autónoma emerge como el punto más frágil.

La valoración global del cuestionario es positiva ($M = 3,86$ sobre 5), aunque el patrón no es homogéneo. La dimensión mejor valorada es Metodologías prácticas ($M = 4,45$), seguida de Tecnología educativa ($M = 3,93$) y Motivación ($M = 3,83$), mientras que Competencia percibida aparece como la dimensión más débil ($M = 3,62$) (véase figura 1). Este resultado indica que el alumnado valora la asignatura y aprecia la ayuda de la tecnología, pero no se siente plenamente seguro al trasladar los conceptos a tareas de análisis y decisión financiera.

Tabla 2. Estadísticos descriptivos por dimensión

Dimensión	Ítems	Media	DE	% con media ≥ 4
Motivación	4	3,83	0,73	57,4%
Metodologías prácticas	4	4,45	0,54	84,0%
Competencia percibida	10	3,62	0,76	33,0%
Tecnología educativa	3	3,93	0,91	60,6%

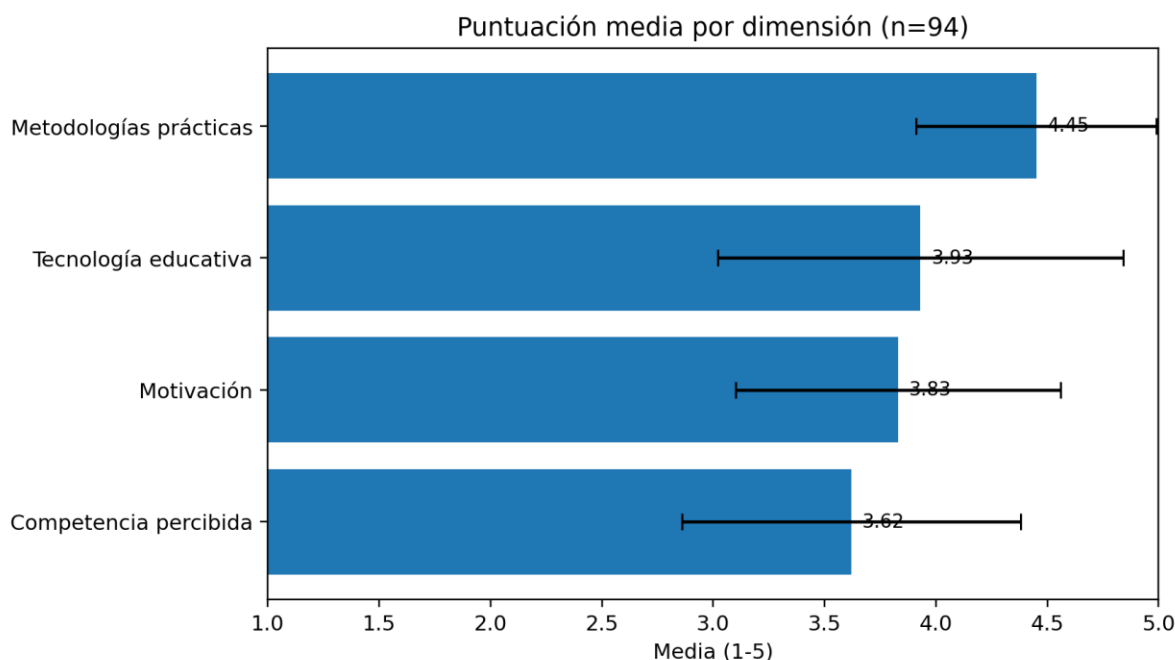


Figura 1. Puntuación media por dimensión en la muestra analizada.

Los ítems con puntuaciones más altas se concentran en el bloque metodológico y muestran un apoyo inequívoco al aprendizaje activo: preferencia por actividades prácticas, utilidad de los ejercicios, valor de los casos reales y ayuda de las simulaciones. En cambio, las puntuaciones más bajas aparecen en tareas de mayor transferencia: identificar información relevante, analizar la situación financiera de una empresa, detectar problemas y utilizar información financiera para apoyar decisiones empresariales (Tabla 3). El ítem con menor media es 'Disfruto resolviendo problemas financieros' ($M = 3,32$), lo que sugiere una brecha entre utilidad percibida y experiencia subjetiva de dominio.

Tabla 3. Selección de ítems con puntuaciones extremas

Ítem	Media	% acuerdo (4-5)
Prefiero actividades prácticas en clase	4,55	92,6%
Los ejercicios prácticos mejoran mi comprensión	4,49	87,2%
Analizar casos reales mejora mi aprendizaje	4,45	89,4%
Las simulaciones ayudan a entender la contabilidad	4,31	83,0%
Aprender contabilidad es importante para mi futuro profesional	4,27	85,1%
Puedo detectar problemas financieros en los estados contables	3,55	51,1%
Puedo analizar la situación financiera de una empresa	3,54	53,2%

Soy capaz de identificar información relevante en los estados financieros	3,53	44,7%
Soy capaz de utilizar información financiera para apoyar decisiones empresariales	3,52	48,9%
Disfruto resolviendo problemas financieros	3,32	50,0%

Las respuestas abiertas (Tabla 4) refuerzan el patrón observado en los datos cuantitativos. La categoría dominante es la solicitud de más ejercicios y casos prácticos (20 menciones; 60,6% de las respuestas abiertas), seguida de la petición de más materiales o soluciones disponibles en el campus virtual (8 menciones; 24,2%) y de más tecnología, inteligencia artificial o software aplicado (6 menciones; 18,2%). Estas demandas no se contradicen entre sí; más bien, dibujan una expectativa integrada de docencia práctica, apoyo continuo y recursos digitales útiles (véase figura 2).

Tabla 4. Temas emergentes en las respuestas abiertas

Tema	Menciones	% sobre respuestas abiertas
Más ejercicios/casos prácticos	20	60,6%
Más material/soluciones en campus	8	24,2%
Más tecnología/IA/software	6	18,2%
Sin cambios / valoración positiva	4	12,1%
Mayor alineación con el examen	4	12,1%
Refuerzo/clases adicionales	2	6,1%
Trabajo colaborativo	2	6,1%

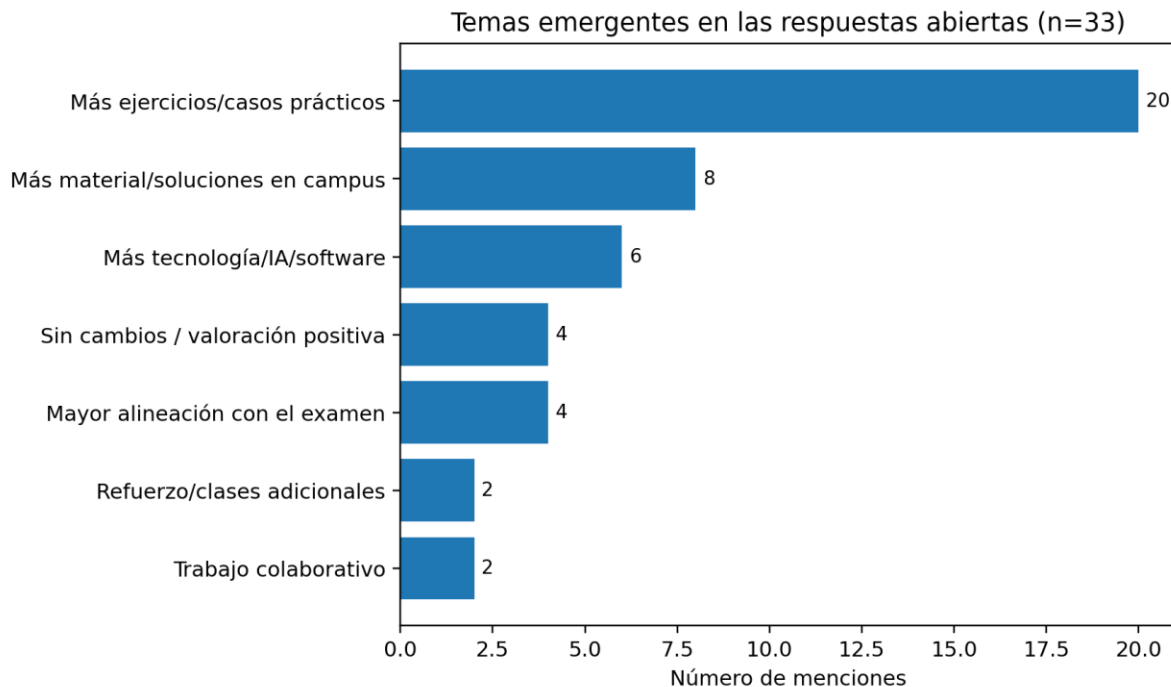


Figura 2. Frecuencia de temas emergentes en la pregunta abierta.

La combinación de resultados cuantitativos y cualitativos permite formular una lectura coherente del problema pedagógico. El alumnado no rechaza la contabilidad; de hecho, reconoce su utilidad para el futuro profesional. Tampoco rechaza el esfuerzo exigente. Lo que aparece con claridad es una necesidad de aprender mediante práctica guiada, casos auténticos, ejemplos resueltos y recursos que faciliten la progresión desde la comprensión inicial hasta la aplicación autónoma.

Este patrón se alinea con la literatura sobre aprendizaje activo y educación contable. La elevada valoración de ejercicios, simulaciones y casos reales sugiere que el alumnado identifica correctamente qué tipos de experiencias le ayudan a comprender. A su vez, las puntuaciones más bajas en competencia percibida indican que todavía no se han consolidado suficientemente las experiencias de dominio necesarias para que el estudiante se sienta capaz de interpretar, diagnosticar y decidir con información financiera.

La cuestión didáctica de fondo, por tanto, no es si debe incorporarse tecnología, sino qué tipo de tecnología y con qué propósito. Una tecnología orientada solo a presentar contenidos difícilmente resolvería la principal debilidad observada. En cambio, una herramienta que aumente el tiempo de práctica, ofrezca feedback inmediato, permita trabajar con escenarios realistas y facilite la intervención temprana del profesorado sí parece alinearse con las necesidades detectadas.

7. Contamind. Herramienta de apoyo a la docencia.

Contamind puede conceptualizarse como un entorno digital de aprendizaje diseñado para responder a una de las principales tensiones actuales de la educación contable: la necesidad de aproximar la formación universitaria a las exigencias del desempeño profesional sin renunciar a la lógica pedagógica propia del aula. La herramienta surge en un contexto caracterizado por una paradoja estructural: elevada empleabilidad de los egresados del ámbito contable y, al mismo

tiempo, percepción de insuficiente adecuación entre la formación universitaria y las demandas reales del mercado laboral (ASEPUC 2024, 2025). Desde esta perspectiva, Contamind se presenta como una propuesta orientada a reducir la brecha entre universidad y empresa mediante la integración de simulación profesional, inteligencia artificial y analítica docente.

A diferencia del software profesional, cuyo objetivo principal es maximizar la eficiencia operativa, automatizar tareas y reducir el error humano, Contamind se inscribe en la categoría de software docente. Esta distinción es relevante desde el punto de vista académico, ya que permite comprender que el valor de la herramienta no reside en ejecutar procesos contables con rapidez, sino en hacer visible el razonamiento contable del estudiante, favorecer la comprensión progresiva y utilizar el error como un recurso pedagógico. En este sentido, el entorno prioriza la secuenciación didáctica, la ayuda contextual y la explicación del error por encima de la automatización ciega, lo que lo convierte en un dispositivo especialmente adecuado para contextos formativos introductorios e intermedios.

La arquitectura de Contamind se articula en torno a tres componentes funcionales. En primer lugar, incorpora un motor de inteligencia artificial que adapta los contenidos al nivel del estudiante, identifica áreas de mejora y ofrece retroalimentación inmediata. En segundo lugar, incluye una capa de simulación profesional que permite trabajar con escenarios contables verosímiles —asientos, regularización, cálculo del resultado, cierre y balances— en un entorno que emula parcialmente la lógica del software empresarial. En tercer lugar, integra un panel de seguimiento docente orientado a la monitorización de consultas, detección de errores recurrentes y evaluación del progreso. Esta triple estructura sugiere que Contamind no debe ser interpretado únicamente como una herramienta de práctica, sino como un sistema de mediación pedagógica que actúa simultáneamente sobre el aprendizaje del estudiante y sobre la capacidad del profesorado para intervenir de forma más informada.

Desde una perspectiva didáctica, la principal aportación de Contamind radica en su potencial para desplazar la enseñanza de la contabilidad desde formatos centrados en la exposición y la memorización hacia modalidades de aprendizaje activo, aplicado y progresivamente autónomo. La herramienta sitúa al alumnado en escenarios de trabajo que requieren ejecutar operaciones, consultar información, revisar errores y contrastar decisiones, aproximando así la experiencia de aprendizaje a los contextos reales de actuación profesional. Esta lógica resulta especialmente pertinente en contabilidad, donde la adquisición de competencias no depende exclusivamente del conocimiento declarativo, sino también de la capacidad para interpretar información financiera, justificar decisiones y transferir procedimientos a situaciones nuevas.

El uso de Contamind por parte del profesorado refuerza esta orientación pedagógica. La plataforma permite gestionar grupos, programar actividades, construir o adaptar ejercicios, introducir instrucciones, fijar tiempos orientativos y asignar tareas diferenciadas. Asimismo, el repositorio de ejercicios posibilita trabajar con distintas configuraciones didácticas, desde ciclos completos hasta ejercicios iniciados desde balances intermedios, lo que amplía la flexibilidad metodológica y facilita la alineación entre práctica, secuenciación del aprendizaje y evaluación. Todo ello sugiere que la herramienta no solo sirve para “hacer ejercicios”, sino para diseñar trayectorias formativas estructuradas en función de objetivos de aprendizaje explícitos.

Desde la perspectiva del estudiante, Contamind organiza el trabajo en torno a tareas disponibles, en progreso o completadas, y pone a su disposición una interfaz centrada en el libro diario, el libro mayor, el balance de comprobación, el plan general contable y distintos apoyos para la revisión del ejercicio. La presencia de un módulo de consulta con IA añade una dimensión de apoyo contextualizado que puede resultar especialmente útil en procesos de práctica autónoma, siempre que dicho apoyo se conciba como andamiaje y no como sustitución del razonamiento. En ese marco, la utilidad de la IA no reside en proporcionar respuestas finales, sino en ofrecer pistas, aclaraciones, ejemplos y retroalimentación que permitan al estudiante avanzar en la resolución del problema con mayor comprensión.

Otro aspecto académicamente relevante es la capacidad de la herramienta para generar trazas de aprendizaje. Variables como número de ejercicios disponibles, progreso, errores, aciertos, consultas realizadas a la IA y tiempo de trabajo aportan información valiosa para el seguimiento docente y para una eventual evaluación basada en procesos. Aunque algunas de estas funcionalidades aparecen todavía en desarrollo, la lógica subyacente es clara: desplazar la evaluación desde una visión exclusivamente sumativa hacia una perspectiva más formativa, en la que los datos de uso permitan detectar dificultades tempranas, ajustar la intervención tutorial y mejorar la alineación entre práctica y evaluación. Desde este punto de vista, Contamind favorece una reconfiguración del rol docente, que pasa de la mera transmisión de contenidos a la facilitación y monitorización de experiencias de aprendizaje.

Con todo, una lectura académica rigurosa exige subrayar que el valor pedagógico de Contamind no depende únicamente de su sofisticación tecnológica, sino de las condiciones de su integración curricular. Una herramienta de estas características solo adquiere sentido cuando se inserta en secuencias didácticas coherentes, complementa la docencia presencial o híbrida y se orienta al fortalecimiento del juicio profesional. En consecuencia, su implantación debe plantearse como un uso complementario, no sustitutivo, acompañado de capacitación docente, evaluación continua y criterios explícitos sobre el papel de la IA en el aprendizaje y en la resolución de tareas.

En la siguiente tabla (Tabla 5) se ha pretendido alinear los resultados obtenidos del diagnóstico del alumnado y la propuesta de valor de Contamind, con el fin de analizar la contribución de la herramienta informática, en base a los usos señalados.

Tabla 5. Alineación entre necesidades detectadas y propuesta pedagógica de Contamind

Necesidad detectada	Funcionalidad de Contamind	Mecanismo pedagógico	Contribución esperada
Predominio de la preferencia por ejercicios, casos y simulaciones (M = 4,45 en metodologías prácticas)	Simulación profesional con tareas contables realistas	Aprendizaje activo y experiencial; transferencia a contextos auténticos (Freeman, et al, 2014; Gittings et al, 2020 y Tran y herzing, 2023)	Mayor comprensión aplicada y cercanía al desempeño profesional
Menor competencia percibida en análisis, diagnóstico y apoyo a decisiones (M = 3,62)	IA adaptativa con pistas, feedback inmediato y práctica escalonada	Andamiaje, refuerzo de autoeficacia y corrección temprana de errores (Honicke y broadbent, 2016; van dinther et al, 2011; Wiseman y Foster, 2025)	Incremento de la confianza operativa y mejora en interpretación financiera
Demanda de más material, soluciones y apoyo fuera del aula	Repositorio dinámico de actividades,	Extensión del tiempo de práctica y aprendizaje autorregulado	Mejor continuidad entre clase, estudio autónomo y evaluación

Necesidad detectada	Funcionalidad de Contamind	Mecanismo pedagógico	Contribución esperada
	ayudas y rutas personalizadas	(DÁquila et al, 2019)	
Petición explícita de más tecnología/IA/software	Entorno integrado de aprendizaje con IA y simulación	Uso significativo de tecnología con propósito pedagógico (Theuri, et al, 2024; Ballantine et al, 2024; Haugland y Kulset, 2024 Zhou y Lou, 2025)	Aceptación estudiantil y experiencia de aprendizaje más rica
Necesidad de detección temprana de dificultades y alineación con la evaluación	Panel docente de seguimiento, errores recurrentes y trazas de uso	Feedback formativo basado en datos e intervención tutorial personalizada (Morris, et al, 2021)	Apoyo focalizado del profesorado y mejor ajuste entre práctica y evaluación

La implantación de Contamind debería organizarse en un modelo de uso complementario, no sustitutivo. La herramienta tendría mayor valor cuando se articula con la asignatura mediante secuencias del tipo: explicación breve del concepto, ejemplo resuelto, práctica en Contamind, retroalimentación automática, discusión de errores frecuentes y transferencia a un caso real o a una prueba de aula.

El motor de IA no debería utilizarse para entregar respuestas finales sin mediación, sino para ofrecer pistas, explicaciones graduadas, preguntas de comprobación y retroalimentación sobre errores. La simulación profesional, por su parte, permitiría acercar la contabilidad al contexto de trabajo con asientos, cierres y balances en un entorno verosímil. El panel docente serviría para detectar estudiantes con dificultades persistentes, tareas donde se concentran errores y momentos del curso que requieren refuerzo.

De este modo, Contamind podría contribuir simultáneamente a tres fines: aumentar la cantidad de práctica significativa, mejorar la calidad del feedback y reforzar la función tutorial del profesorado. Estos tres objetivos son especialmente pertinentes en contextos de primer y segundo curso, donde la heterogeneidad de entrada suele ser elevada y la autorregulación académica todavía se está consolidando.

8. Conclusiones y líneas futuras de investigación

Aunque este trabajo proporciona una justificación académica razonada, la eficacia de Contamind debe verificarse empíricamente. La evaluación futura más adecuada sería un diseño cuasi experimental con grupo de comparación, medidas pretest-postest y combinación de indicadores de percepción, uso y rendimiento.

Entre los indicadores prioritarios deberían incluirse: (a) motivación hacia la asignatura, (b) autoeficacia o competencia percibida en análisis e interpretación financiera, (c) rendimiento en pruebas auténticas de resolución de casos, (d) trazas de uso de la plataforma: consultas, errores recurrentes, tiempo de práctica y (e) satisfacción estudiantil y docente.

Además, resultaría especialmente relevante analizar si el efecto de la herramienta es mayor en estudiantes con menor competencia inicial, así como estudiar qué patrones de interacción con la IA se asocian con mejores resultados. Esta línea permitiría pasar de la justificación pedagógica a la validación empírica del diseño.

Tabla 6. Indicadores sugeridos para evaluar Contamind en futuras investigaciones

Dimensión a evaluar	Instrumento o fuente	Resultado esperado
Motivación hacia la contabilidad	Cuestionario pretest-postest	Incremento medio en motivación y disfrute
Competencia percibida / autoeficacia	Escala Likert por tareas financieras	Mejora en análisis, interpretación y decisión
Rendimiento académico	Pruebas auténticas y casos evaluables	Mejora en resultados respecto a grupo comparación
Uso de la herramienta	Logs de consultas, errores, tiempo y progresión	Patrones de uso asociados a mejor aprendizaje
Calidad del razonamiento	Rúbricas de justificación y revisión crítica	Mayor juicio profesional y menor dependencia acrítica

La principal aportación de este trabajo es mostrar que la justificación del uso de Contamind puede formularse en términos académicos sólidos y no solo en clave de transformación digital. Los datos del alumnado indican que la mejora del aprendizaje de la contabilidad pasa por más práctica, más casos, más apoyo estructurado y una integración significativa de la tecnología.

Contamind resulta pedagógicamente pertinente porque sus tres componentes -IA adaptativa, simulación profesional y panel docente- se alinean con esas necesidades y con la evidencia disponible sobre aprendizaje activo, autoeficacia, feedback formativo y educación contable apoyada en tecnología. Ahora bien, dicha pertinencia depende de un principio fundamental: la herramienta debe fortalecer el razonamiento contable y el juicio profesional, no reemplazarlos.

Referencias

- ASEPUC (2024), Encuesta a profesores universitarios sobre la formación de los estudiantes en contabilidad y auditoría (Coord.: García Meca, E. y López Espinosa, G.). En <https://asepuc.org/zona-descarga/encuesta-a-profesoresuniversitarios-sobre-laformacion-de-losestudiantes-encontabilidad-yauditoria/>
- ASPEPUC (2025), La profesión contable y auditora en España: Formación, demanda del mercado laboral y recomendaciones. (Coord.: García Meca, E. y Martínez Ferrero, J.). En <https://asepuc.org/zona-descarga/encuesta-a-responsables-de-departamentos-de-contabilidad-gerentes-socios-y-otros-perfiles-claves/>
- Ballantine, J., Boyce, G., & Stoner, G. (2024). A critical review of AI in accounting education: Threat and opportunity. *Critical Perspectives on Accounting*, 99, 102711. <https://doi.org/10.1016/j.cpa.2024.102711>

- D'Aquila, J. M., Wang, D., & Mattia, A. (2019). Are instructor generated YouTube videos effective in accounting classes? A study of student performance, engagement, motivation, and perception. *Journal of Accounting Education*, 47, 63-74. <https://doi.org/10.1016/j.jaccedu.2019.02.002>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415. <https://doi.org/10.1073/pnas.1319030111>
- Gittings, L., Taplin, R., & Kerr, R. (2020). Experiential learning activities in university accounting education: A systematic literature review. *Journal of Accounting Education*, 52, 100680. <https://doi.org/10.1016/j.jaccedu.2020.100680>
- Haugland Sundkvist, C., & Kulset, E. M. (2024). Teaching accounting in the era of ChatGPT - The student perspective. *Journal of Accounting Education*, 69, 100932. <https://doi.org/10.1016/j.jaccedu.2024.100932>
- Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review*, 17, 63-84. <https://doi.org/10.1016/j.edurev.2015.11.002>
- Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, 9(3), e3292. doi:10.1002/rev3.3292.
- Theuri, P., Campbell, R., & Owens-Jackson, L. (2024). A literature review of technology-related research in accounting education: 2010-2020. *Accounting Perspectives*, 23(1), 79-114. <https://doi.org/10.1111/1911-3838.12352>
- Tran, T. T., & Herzig, C. (2023). Blended case-based learning in a sustainability accounting course: An analysis of student perspectives. *Journal of Accounting Education*, 63, 100842. <https://doi.org/10.1016/j.jaccedu.2023.100842>
- van Dinther, M., Dochy, F., & Segers, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review*, 6(2), 95-108. <https://doi.org/10.1016/j.edurev.2010.10.003>
- Wiseman, D., & Foster, D. (2025). Bridging AI with cost accounting education: A pedagogical framework using reperformance and critical error review. *Journal of Accounting Education*, 72, 100988. <https://doi.org/10.1016/j.jaccedu.2025.100988>
- Zhou, A., & Luo, Y. (2025). Exploring the impact of generative AI on student learning in accounting. *Journal of Accounting Education*, 72, 100982. <https://doi.org/10.1016/j.jaccedu.2025.100982>

ContaMIND: an integrated IT architecture for automated formative feedback, learning analytics, and controlled RAG tutoring in accounting education

José Miguel Mota^{1*} Rubén Baena Pérez¹
Paula Rodríguez Castro² Jesús Rodríguez Torrejon²
Andrés Muñoz Ortenga¹

¹Departamento de Ingeniería Informática, Escuela Superior de Ingeniería, Universidad de Cádiz, Puerto Real, Spain.

²Departamento Economía Financiera y Contabilidad, Facultad de CC.EE. y Empresariales, Universidad de Cádiz, Cádiz, Spain.

Journal: Education and Information Technologies (Springer, Q1)

Software registration: ContaMIND is registered by Universidad de Cádiz as *Programa de ordenador*, file CA-2605085574200-2026, concession date 08/05/2026.

Teaching innovation project: *Desarrollo de Software Docente para la Enseñanza en Contabilidad Financiera: Spreadify*, ref. sol-202300256638-tra.

Data governance: Data-processing procedure notified to the Universidad de Cádiz Data Protection Officer (dpd@uca.es). All analyses use fully anonymised data; no direct personal identifiers are present in the analytical dataset.

Conflicts of interest: The authors declare no conflicts of interest.

Keywords: learning analytics • automated formative feedback • educational data mining • accounting education • teacher-facing dashboards • interaction traces • information technology in higher education • controlled retrieval-augmented generation

*Corresponding author. E-mail: josemiguel.mota@uca.es

Iterative Practice in Accounting Education: Deployment Evidence from the ContaMIND Learning Analytics and Automated Feedback Architecture

Author 1^{a,*}, Author 2^a, Author 3^a

^aDepartamento de Economía Financiera y Contabilidad, Facultad de Ciencias Sociales y de la Comunicación, Universidad de Cádiz, Jerez de la Frontera, Spain

Abstract

Accounting education requires extensive procedural practice, formative feedback, and opportunities to transfer technical knowledge to professional scenarios. General-purpose generative AI poses specific risks in this domain, including the creation of hallucinated accounting entries and limited teacher visibility into student learning. This paper presents ContaMIND, an AI-supported accounting learning environment integrating four layers: a professional simulation where students construct journal entries; an automatic review engine that evaluates submissions against instructor-defined solutions at the entry and field level; a controlled retrieval-augmented generation (RAG) tutor constrained to instructor-curated course materials; and a teacher-facing learning analytics dashboard with configurable alert-rule support specified as a future-cycle design module. The design responds to an empirically identified need: a prior survey of 94 students revealed a strong preference for practical activities ($M = 4.45/5$) alongside limited perceived competence in financial analysis ($M = 3.62/5$). The system was deployed in a compulsory first-year financial accounting course, involving 194 unique student accounts across two undergraduate programmes. The deployment dataset comprises 464 exercise attempts, 9,735 automatic reviews, 390,828 field-level change records, and 153 AI-tutor queries. An analysis of reviewed attempts shows a mean first-review score of 11.29% increasing to 69.45% at last review; we interpret this within-platform improvement as a process indicator of sustained feedback engagement, not as a causal learning outcome. This interpretation is supported architecturally by an instructor-defined `num_verificacion` state-machine gate that disables further review until the student opens the controlled RAG tutor. An internal-balance diagnostic reveals that 121,461 entry-level review observations were structurally balanced but content-incorrect, illustrating the value of fine-grained formative diagnostics beyond simple correctness scores. We present ContaMIND as a controlled, auditable architecture for AI-supported accounting education; empirical evaluation of the RAG tutor and future calibration of the alert-rule module are scoped as future work.

Keywords: artificial intelligence in education, retrieval-augmented generation, learning analytics, accounting education, formative feedback, intelligent tutoring systems, higher education

*Corresponding author

Email address: (Author 1)

1. Introduction

Accounting education in higher education faces a persistent pedagogical challenge: students must learn technical rules and procedures while also developing the ability to interpret financial information, justify accounting decisions, and diagnose errors in realistic business scenarios. Prior research in active learning and accounting education suggests that experiential activities, cases, and applied tasks improve understanding and engagement when aligned with clear learning objectives (Freeman et al., 2014; Gittings et al., 2020; Tran and Herzig, 2023).

This challenge is well documented in the local context motivating the present study. A prior survey of 94 first- and second-year university students in accounting courses revealed an overall positive evaluation of accounting learning ($M = 3.86/5$), with the highest scores concentrated in the methodological dimension: preference for practical activities ($M = 4.45/5$, 84.0% agreement) and utility of real cases and simulations. In contrast, perceived competence in financial analysis and interpretation was the weakest dimension ($M = 3.62/5$, 33.0% reaching the threshold of agreement), and open-ended responses showed that 60.6% of students requested more practical exercises and cases, while 18.2% explicitly requested more technology, AI, or educational software (Anonymous, 2026). This gap between engagement with practice and confidence in professional competence defines the pedagogical problem that ContaMIND is designed to address.

At the same time, generative AI is changing the conditions under which students seek support. General-purpose tools can explain concepts and produce worked examples, but they may also generate unsupported or incorrect accounting entries. In educational settings, this creates a tension between the potential of AI as a scaffold and the risk that students may outsource reasoning to an opaque system. Recent work in AI and accounting education has emphasised both the opportunities and risks of generative AI, particularly the need to preserve professional judgement, critical error review, and instructor oversight (Ballantine et al., 2024; Haugland Sundkvist and Kulset, 2024; Wiseman and Foster, 2025; Zhou and Luo, 2025).

This paper presents ContaMIND, a web-based learning environment designed for university accounting courses. The system integrates four layers: (1) an accounting simulation environment where students solve exercises using journal entries; (2) an automatic review engine that compares student work with instructor-defined solutions at entry and field level; (3) a controlled Retrieval-Augmented Generation (RAG) tutor connected to curated course materials through a deterministic n8n workflow; and (4) a teacher dashboard that transforms interaction traces into learning analytics indicators, with configurable alert rules implemented as a dormant module for future educational cycles. The intended contribution is the design of an auditable educational architecture in which AI support, student practice, automatic formative feedback, and teacher analytics are technically integrated.

The platform was piloted during the second semester of the 2025/2026 academic year in the subject *Fundamentos de Contabilidad Financiera* (Fundamentals of Financial Accounting), a compulsory first-year course delivered face-to-face at the Faculty of Social Sciences, Jerez de la Frontera campus, to students enrolled in two undergraduate programmes: the Degree in Public Management and Administration and the Degree in Marketing and Advertising. The implementation involved 197 active enrolments (57 in Public Management and Administration;

140 in Marketing and Advertising), with 194 unique student accounts after removing three
accounts appearing in both groups. The platform contributed 20% of the final course grade.
The target-group activity window runs from 6 April 2026 to 21 May 2026. ContaMIND
originated from the teaching innovation project *Desarrollo de Software Docente para la*
Enseñanza en Contabilidad Financiera: Spreadify (ref. sol-202300256638-tra); the software
was developed under the name Spreadify and subsequently registered by the University of
Cádiz under the name ContaMIND (file CA-2605085574200-2026).

The paper addresses the following research questions:

RQ1. How do students interact with an AI-supported accounting learning environment
during real practice activities?

RQ2. What patterns of revision, correction, persistence, and automatic feedback use are
observable in student interaction traces during accounting practice?

RQ3. What types of accounting difficulties emerge from student traces, automatic reviews,
and AI-tutor queries?

RQ4. How is the environment’s architecture designed to support teacher visibility over
problematic exercises and students who may require additional formative support?

Research Questions (RQ) 1–4 are addressed using the deployment learning-trace dataset.
Technical evaluation of the RAG tutor against an expert-labelled benchmark is planned as
future work following completion of the benchmark design.

The main contributions of this paper are: (1) a formally specified, deterministic automatic
review algorithm for accounting exercises, with a functional verification framework specifying
nine behavioural conditions at field, entry, and attempt levels — the algorithm and its
deployment evidence constitute the primary empirical contribution; (2) a real deployment
dataset from 194 students comprising 464 attempts, 390,828 field-level change records, and
9,735 automatic reviews, analysed as process indicators of feedback engagement and accounting
difficulty patterns; (3) a client-side state-machine gatekeeper that uses the instructor-defined
`num_verificacion` ceiling to interrupt unrestricted review requests and route students through
the controlled RAG tutor before additional checks are unlocked; (4) a teacher-facing learning
analytics dashboard with a modular alert-rule specification, intentionally left unconfigured
during the initial deployment to avoid instructor alert fatigue; and (5) as an architectural
contribution only, a controlled RAG tutoring pipeline implemented through a deterministic n8n
workflow with semantic query expansion, constrained generation, and documented safeguards
— formal empirical evaluation of the tutor is scoped as future work.

The remainder of the paper is organised as follows. Section 2 reviews related work.
Section 3 describes the technological architecture. Section 4 describes the methodology,
including study design, participants, data collection, analytical indicators, and statistical
procedures. Section 5 reports results. Section 6 discusses the implications, limitations, and
future work.

Although ContaMIND is implemented for financial accounting, the architectural contri-
bution is relevant to AI-supported education more broadly. The integration of controlled
generative AI, deterministic formative assessment, fine-grained learning traces, and teacher-
facing analytics addresses a general challenge in technology-enhanced learning: how to provide

AI scaffolding without compromising instructor oversight or student reasoning. The design principles are transferable to other procedural domains where reference solutions, stepwise production, and instructor oversight are central, such as bookkeeping, auditing, or tax preparation exercises.

2. Background and Related Work

2.1. Active learning, self-efficacy, and accounting education

Active learning has been associated with improved student performance in higher education, especially when learners engage with problems, cases, and iterative practice rather than receiving only expository instruction (Freeman et al., 2014). In accounting education, systematic reviews have highlighted the value of experiential learning activities and case-based learning for connecting technical knowledge with authentic professional tasks (Gittings et al., 2020; Tran and Herzig, 2023). The accounting education research literature has been periodically synthesised in comprehensive reviews (Apostolou et al., 2018), which identify technology integration, active learning, and assessment innovation as persistent themes. These findings support the use of simulation environments where students can repeatedly construct, review, and revise accounting entries.

From a learning design perspective, accounting tasks involve multiple cognitive levels. At the foundational level, students must recall account codes, classification rules, and Value Added Tax (VAT) treatment; at higher levels, they must analyse transactions, evaluate alternative treatments, and construct entries that reflect professional judgement. This progression aligns with Bloom’s revised taxonomy (Kratwohl, 2002), and suggests that automated feedback systems should support not only recall-level verification (correct/incorrect account code) but also the diagnostic information needed to scaffold higher-order reasoning. The concept of “threshold concepts” — transformative ideas that, once understood, irreversibly change how a learner views a domain (Meyer and Land, 2003) — has been applied to accounting to identify concepts such as duality (double-entry), accruals, and materiality as thresholds that students must cross to achieve competent practice (Lucas and Mladenovic, 2006). The balance-versus-correctness diagnostic in ContaMIND is directly relevant to the duality threshold: a student who consistently produces balanced but content-incorrect entries may have grasped the structural principle of double entry without yet understanding which accounts to use for specific transactions.

Beyond achievement, self-efficacy is a key mechanism through which practice environments support learning. A systematic review of 59 studies found a moderate positive relationship between academic self-efficacy and performance (Honicke and Broadbent, 2016). Programmes incorporating mastery experiences, timely feedback, and instructional scaffolding have been shown to strengthen self-efficacy (Van Dinther et al., 2011). In accounting specifically, mastery feedback in introductory courses has been associated with stronger self-efficacy beliefs (Beatson et al., 2018). The gap between engagement with practice ($M = 4.45$) and perceived professional competence ($M = 3.62$) confirmed in the target population (Anonymous, 2026) illustrates this pattern directly.

The role of technology in supporting accounting education has also been reviewed systematically. A literature review of 187 articles on technology-related research in accounting

education from 2010 to 2020 identified technology integration as a consolidated axis of curricular innovation, including the use of simulation tools, digital feedback, and online learning environments (Theuri et al., 2024). Earlier empirical work found that instructor-generated videos improved student performance and were perceived as useful supplementary resources, though not as substitutes for in-class instruction (D’Aquila et al., 2019).

Formative feedback is another central mechanism. It can help students identify gaps between current and expected performance, particularly when it is timely, specific, and connected to actionable next steps (Morris et al., 2021; Shute, 2008). In accounting tasks, feedback is especially important because errors may arise from incorrect account selection, wrong debit/credit allocation, amount miscalculation, missing VAT treatment, or misunderstanding of the underlying transaction. A digital environment can capture these errors at fine granularity and transform them into analytics for both students and teachers.

ContaMIND responds directly to the needs identified in this literature: students benefit from mastery experiences (repeated practice with structured feedback), scaffolded support that preserves rather than replaces their reasoning, and timely teacher monitoring to intervene before difficulties become entrenched.

2.2. Generative AI, RAG, and controlled educational support

The growing use of generative AI in education has led to renewed interest in how AI systems can support students without replacing disciplinary reasoning. In accounting education, recent contributions have warned against purely instrumental uses of AI and proposed pedagogical designs based on reperformance, critical error review, and reflective use of AI-generated outputs, including in the context of emerging professional competency frameworks that explicitly address generative AI skills (Ballantine et al., 2024; Wiseman and Foster, 2025; Anica-Popa et al., 2024). Retrieval-augmented generation (RAG) provides a technical strategy for constraining generative responses to a curated knowledge base, although its educational value depends on how retrieval, prompting, refusal behaviour, logging, and teacher oversight are implemented.

In educational applications, RAG has been proposed as a strategy for providing contextualised, curriculum-aligned generative support while limiting exposure to out-of-domain or fabricated content (Li et al., 2025). RAG systems reduce, but do not eliminate, hallucinations, and therefore require specific evaluation procedures that examine both the retrieval and generation components. Current RAG evaluation frameworks distinguish between retrieval relevance, answer relevance, factual correctness, faithfulness with respect to retrieved context, and the ability to refuse unsupported answers (Lewis et al., 2020; Es et al., 2024; Niu et al., 2024; Yu et al., 2024). This requirement is especially important in educational domains, where inaccurate feedback may mislead learners. Recent work on Large Language Model (LLM)-based formative feedback and intelligent tutoring systems stresses the need for theoretically grounded feedback, empirical evaluation, and integration with learning analytics rather than direct substitution of human teaching (VanLehn, 2011; Stamper et al., 2024; Yan et al., 2024; Cukurova, 2024).

ContaMIND positions itself in this literature by designing the RAG tutor as a controlled, deterministic pipeline: responses are constrained to instructor-curated materials, unsupported questions trigger a fixed refusal message, generation is deterministic (temperature = 0), and all interactions are logged for audit. The knowledge base is populated exclusively by instructors

through a dedicated document upload workflow, ensuring that the RAG system can only draw on content that has been deliberately selected and made available for the course. The distinction between this architecture and a general-purpose AI chatbot is not merely technical; it reflects a pedagogical commitment to preserving student reasoning while providing targeted support.

2.3. Automated assessment and fine-grained formative feedback

Automated assessment in technology-enhanced learning has been studied across multiple domains. In these contexts, automated feedback systems provide timely, scalable formative assessment that would otherwise require excessive instructor effort (VanLehn, 2011). The effectiveness of automated feedback depends not only on accuracy but also on granularity: feedback targeting specific errors rather than aggregate performance has been associated with greater learning gains (Shute, 2008; Stamper et al., 2024).

In procedural domains such as accounting, automated assessment must address the multi-component structure of the task. Accounting journal entries involve account code selection, debit/credit allocation, amount calculation, and VAT treatment — each independently correct or incorrect. A system evaluating only the final grade loses the diagnostic information needed to guide iterative correction. Fine-grained automated feedback can instead identify whether an error arises from incorrect account selection, wrong debit/credit assignment, or a missing line, enabling targeted formative support.

The review mechanism in ContaMIND operates at the entry level rather than the raw interface-row level. Each journal entry is evaluated as a unit, comparing the student’s aggregated debit and credit amounts — and the account codes contributing to them — against the instructor-defined solution. This design allows the system to separate balance errors from content errors: it distinguishes structurally balanced but content-incorrect entries from entries that are both unbalanced and content-incorrect. The diagnostic richness of this output underlies the learning analytics indicators described in subsequent sections.

Automated accounting review also raises a challenge not present in domains such as programming: the existence of pedagogically equivalent alternative solutions. In accounting, multiple account codes may be acceptable for the same transaction depending on the pedagogical framework or the version of the general accounting plan being applied. The current implementation uses exact account-code matching, which is deterministic and auditable but may penalise correct alternatives. This limitation is discussed explicitly in Section 6.

2.4. Learning analytics and teacher-facing dashboards

Learning analytics involves the collection, analysis, and reporting of learner data to understand and optimise learning processes (Siemens and Gasevic, 2011; Long and Siemens, 2011). Teacher-facing dashboards can make student progress, engagement, and difficulty patterns visible, but their usefulness depends on whether indicators are interpretable and actionable. Previous research has shown that many systems remain limited to descriptive indicators, whereas more effective dashboards help users interpret data and decide on next actions (Susnjak et al., 2022; Dourado et al., 2021; Kaliisa et al., 2023). In teacher-facing contexts, dashboard adoption depends on alignment between analytics indicators, pedagogical needs, and teachers’ capacity to act on the information provided (Kaliisa and Dolonen, 2023). This is particularly relevant in higher education, where dashboards are used to monitor

interaction patterns, identify students requiring additional formative support, and guide feedback (Wong and Li, 2020; Romao et al., 2023).

From an educational data mining perspective, platform logs from accounting exercises, revisions, edits, and AI-tutor queries can be transformed into indicators of practice intensity, persistence, recurrent errors, task difficulty, and progression over time (Romero and Ventura, 2020). The ContaMIND dashboard is designed as an actionable analytics layer: it aggregates process data and translates them into teacher-oriented indicators for early detection of difficulties, targeted feedback, and evidence-informed instructional adjustment. In addition to these deployed indicators, the architecture includes a configurable alert-rule module that can be populated in future cycles to identify students who may benefit from additional formative support, based on indicators such as persistent difficulty, lack of progress between reviews, or atypical AI-support use patterns.

2.5. Positioning relative to existing systems

While the individual components of ContaMIND draw on established techniques, our review did not identify a prior system that combines all four layers — simulation, automatic review, controlled AI tutoring, and configurable teacher analytics — within a single auditable architecture for accounting education. Table 1 summarises how ContaMIND relates to existing approaches across adjacent domains.

The table illustrates that existing systems typically address one or two of the four layers. The contribution of ContaMIND is not any single component but their integration into a coherent, auditable architecture designed for procedural accounting practice, combining entry-level diagnostic feedback with controlled AI support and configurable teacher analytics. The scope of this comparison is limited to systems identified in the related literature; the absence of a system from this table does not preclude the existence of comparable approaches.

3. System Overview and Technological Architecture

ContaMIND is a web-based learning environment integrating four tightly coupled components: an accounting simulation interface where students construct and edit journal entries; an automatic review engine that compares submissions with instructor-defined solutions at the entry and field level; a controlled RAG tutor constrained to curated course materials; and a teacher-facing learning analytics dashboard with a configurable alert-rule module reserved for future educational cycles. This section describes the design goals that shaped the architecture, the high-level technical structure, and each component in detail.

3.1. Design goals

ContaMIND was designed around five socio-technical goals derived from both the empirical needs identified in the target student population and the theoretical mechanisms reviewed in the background:

- G1. **Guided accounting practice:** provide students with a digital environment for solving accounting exercises through structured journal entries, emulating the logic of professional accounting software.

Table 1: Positioning of ContaMIND relative to existing approaches.

Approach	Domain practice	Auto. review	AI tutor	Teacher analytics	Key distinction
Programming intelligent tutoring systems			Partial	Partial	Review via test cases; no entry-level accounting diagnostics.
Mathematics intelligent tutoring systems			Adaptive		Adaptive item selection; no procedural journal-entry structure.
General-purpose LLM chatbots	—	—	Uncontrolled	—	No curated retrieval, no refusal policy, no teacher visibility.
RAG-based educational tools	—	—		Partial	Focus on retrieval quality; typically no domain-specific review or simulation.
Accounting textbook platforms		Partial	—	Basic	Multiple-choice or fill-in-the-blank; no entry-level formative diagnostics.
ContaMIND			Controlled RAG		Integrates all four layers; entry-level diagnostic review; deterministic RAG with curated knowledge base; modular alert-rule specification.

- G2. **Automatic formative review:** compare student submissions with instructor-defined solutions and generate entry-level and field-level diagnostics to support iterative correction.
- G3. **Controlled AI support:** offer a RAG-based tutor constrained to course materials, preserving student reasoning rather than replacing it.
- G4. **Traceability:** log student actions, revisions, edits, AI interactions, and temporal activity to enable process-oriented learning analytics.
- G5. **Teacher analytics:** transform interaction traces into interpretable indicators for monitoring groups, exercises, and individual students, with configurable alert rules available as a modular extension.

These goals address the three mechanisms identified as central to improving accounting learning outcomes: increasing the volume of structured practice (G1, G2), providing scaffolded support that strengthens self-efficacy (G3), and enabling data-informed teacher intervention (G4, G5) (Anonymous, 2026; Van Dinther et al., 2011; Morris et al., 2021).

Figures 1 and 2 illustrate the two main analytics interfaces of ContaMIND. The teacher-facing dashboard provides aggregated indicators for monitoring group activity, exercise progression, grade distributions, problematic tasks, and low-progress conditions. The student-facing dashboard presents individual progress indicators computed from the student’s own interaction traces: number of scheduled, started, and completed exercises; mean grade across closed exercises; comparison with the group average (including percentile rank and delta); total time spent; AI-tutor query count; per-exercise progress with grade and group comparison charts; a temporal evolution series of review scores; and formative diagnostics identifying strengths (best exercise), weaknesses (worst exercise), exercises not yet reviewed, and exercises still in progress.

3.2. High-level architecture

Figure 3 summarises the architecture. The application combines a browser-based interface, PHP: Hypertext Preprocessor (PHP) endpoints, MySQL persistence, a review engine, a teacher dashboard, and an n8n-mediated RAG workflow that connects Redis, Qdrant, Ollama, and MySQL logging.

3.3. Application layer

The client side is a web application (HyperText Markup Language (HTML), Cascading Style Sheets (CSS), JavaScript (JS)) with views for accounting journals, exercises, dashboards, and configuration. The backend comprises 14 PHP endpoints covering authentication, exercise management, task submission and review, teacher and student dashboards, alert-rule configuration, AI-tutor chat, and document uploads. The full endpoint listing is included in the reproducibility package (Appendix Appendix B).

The operational user model differentiates between student, teacher, and administrator roles (194 student accounts, 8 teacher, 1 administrator; 209 total). Teacher and administrator accounts are excluded from student learning analytics.

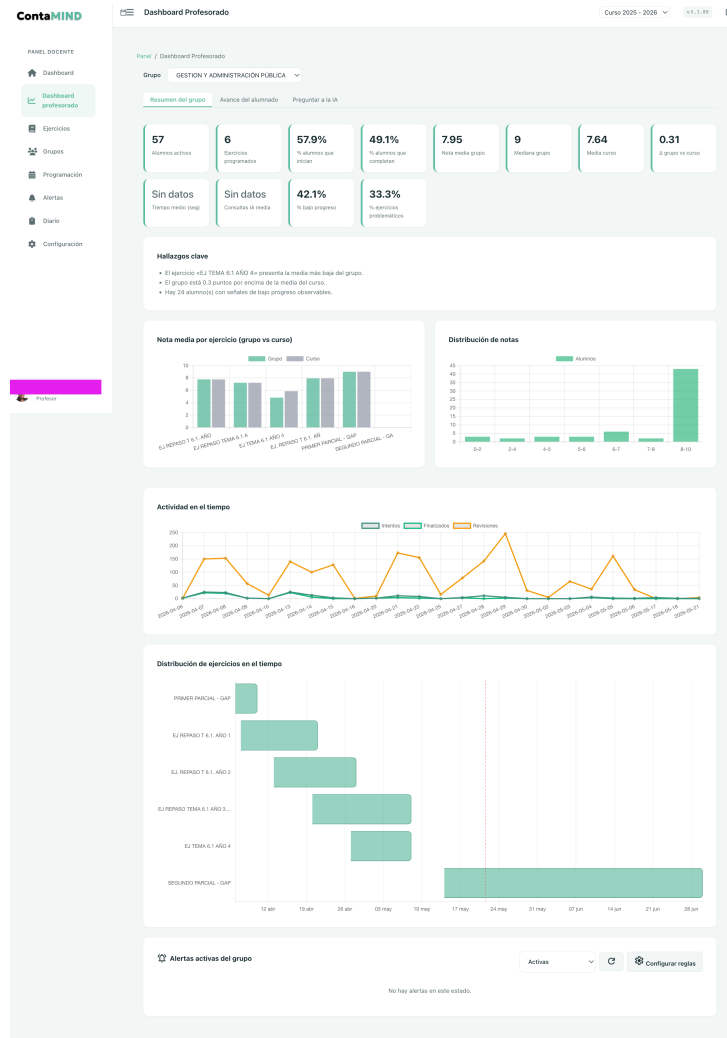


Figure 1: Teacher-facing dashboard showing group-level learning analytics. The dashboard aggregates indicators such as active students, scheduled exercises, initiation and completion rates, group performance, grade distribution, temporal activity, exercise availability windows, and low-progress signals. All screenshots were anonymised before submission.

Dashboard de progreso

Resumen de tus ejercicios, notas y comparativa con tu grupo.



Comparativa con la media del grupo



Fortalezas y puntos de mejora



Detalle por ejercicio

Ejercicio	Estado	Nota	% actual
EJ. TEMA 6.1 AÑO 4	En progreso	=	=
EJ. REPASO TEMA 6.1 AÑO 3 SERV. GAP	Sin iniciar	=	=
EJ. REPASO T. 6.1. AÑO 2	Sin iniciar	=	=
EJ. REPASO T. 6.1. AÑO 1	Sin iniciar	=	=
PRIMER PARCIAL - GAP	Sin iniciar	=	=

Figure 2: Student-facing progress dashboard showing individual exercise status, grade for each completed exercise, comparison with the group average (including percentile rank and delta), revision history evolution, AI-tutor query count, and formative diagnostics (best/worst exercise, exercises in progress, exercises not yet reviewed). All screenshots were anonymised before submission.

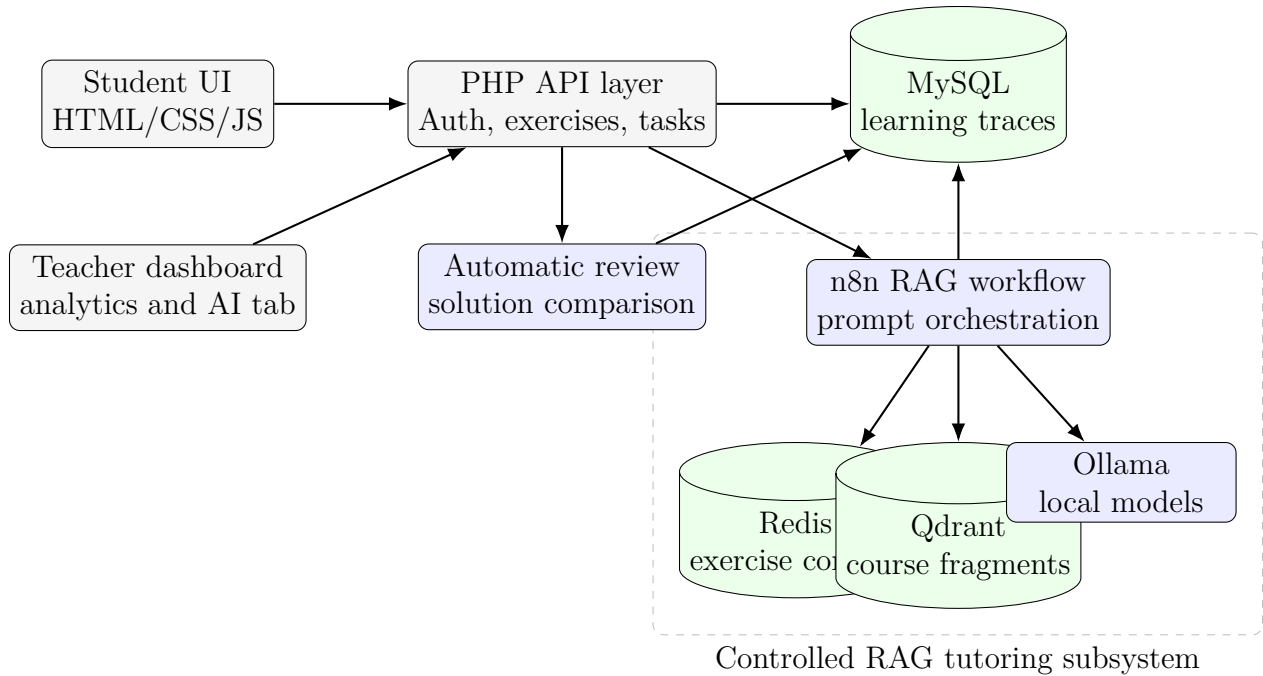


Figure 3: High-level technological architecture of ContaMIND.

3.4. Deployment context and operational security

ContaMIND was deployed at <https://sibila.uca.es/contamind> on a dedicated university research server. AI-related workflows are integrated through webhook-based n8n services; Redis, Qdrant, and Ollama run on the same server. User authentication uses JSON Web Token (JWT) with role-based access control (three roles: Admin, Prof, Alum), enforced at each endpoint. Group-level permissions ensure instructors access only authorised groups. Communication uses HTTPS with standard security headers. Credentials and secrets are stored as server-side configuration constants and excluded from the source-code package.

3.5. Reproducibility and source-code availability

The reproducibility package (Appendix [Appendix B](#)) supports both architectural reproducibility (inspecting and redeploying the system) and empirical reproducibility (reproducing the reported analyses from anonymised data). The RAG evaluation benchmark will be included once expert labelling is complete. Raw database exports are not released because they contain personally identifiable information; the anonymised analytical dataset retains only the variables required to reproduce the statistical analyses (see Section [4.5](#)).

3.6. Accounting simulation and automatic review

Exercises are defined by instructors and stored with statements, metadata, opening and closing dates, publication status, balance/regularisation/closing flags, and a configurable review limit (`num_verificacion`). Figure [4](#) shows the exercise listing view, from which students access available tasks within their group.

ContaMIND

Ejercicios

Curso 2025 - 2026

15.3.26

PANEL DOCENTE

Dashboard

Dashboard profesorado

Ejercicios

Grupos

Programación

Alumnos

Diario

Configuración

Problemas

Repositorio de Ejercicios

Gestiona tu biblioteca de ejercicios contables. Crea nuevos problemas, organiza por temática y supervisa el progreso de los alumnos.

+ Nuevo Ejercicio

Buscar ejercicio por título o código...

Todas las Temáticas

Todos

Todos los permisos

MM

SEGUNDO PARCIAL MM 2026

MARKETING CREATIVO S.A. presenta el siguiente asiento de apertura del AÑO...

COMPLETADO

NOTA MEDIA

Publicado

Hace 1 año

PROPIETARIO

GAP

SEGUNDO PARCIAL - GAP

SERVICIOS GAP S.A. presenta el siguiente asiento de apertura del AÑO...

COMPLETADO

NOTA MEDIA

Publicado

Hace 2 años

PROPIETARIO

MM

EJ. 12 MARKETING CREATIVO AÑO 4

MARKETING CREATIVO AÑO 40.Barbap.Barbap.Barbap.Barbap...

COMPLETADO

NOTA MEDIA

Publicado

Hace 3 años

PROPIETARIO

GAP

EJ.TEMA 6.1 AÑO 4

SERVICIOS GAP S.A. AÑO 4 Balance de...

COMPLETADO

NOTA MEDIA

Publicado

Hace 1 mes

PROPIETARIO

MM

EJERCICIO 11 MARKETING CREATIVO AÑO 3

SERVICIOS 112 MARKETING CREATIVO S.A. Los enunciados en abul...

COMPLETADO

NOTA MEDIA

Publicado

Hace 11 meses

PROPIETARIO

GAP

EJ REPASO TEMA 6.1 AÑO 3 SERV. GAP

112.Barbap.Barbap.Barbap.Barbap de apertura. 1 A partir del asiento de cierre...

COMPLETADO

NOTA MEDIA

Publicado

Hace 1 mes

PROPIETARIO

SALARIOS

AJUSTES CIERRE

MM

EJERCICIO 10 MARKETING CREATIVO AÑO 2

MARKETING CREATIVO S.A. AÑO 20.Barbap.Barbap.Barbap.Barbap...

COMPLETADO

NOTA MEDIA

Publicado

Hace 1 mes

PROPIETARIO

COMPAS

IVA

SALARIOS

GAP

EJ. REPASO T 6.1. AÑO 2

BALANCE DE SITUACION INICIAL SERVICIOS GAP S.A. AÑO 20.Barbap...

COMPLETADO

NOTA MEDIA

Publicado

Hace 1 mes

PROPIETARIO

CONTAS

DEBITOS

IVA

GAP

EJ REPASO T 6.1. AÑO 1

La empresa objeto de nuestra contabilidad es SERVICIOS GAP S.A. e...

COMPLETADO

NOTA MEDIA

Publicado

Hace 11 meses

PROPIETARIO

DEBITOS

IVA

ANULIZACION

MM

EJERCICIO 9 TEMA 8 MARKETING CREATIVO AÑO 1

AÑO 1 MARKETING CREATIVO S.A. es una agencia de publicidad y...

COMPLETADO

NOTA MEDIA

Publicado

Hace 1 mes

PROPIETARIO

DEBITOS

IVA

IVA

IVA

EJERCICIO 8 TEMA 8

Realizar los asientos desde la constitución hasta el cierre.Barbap,Se...

COMPLETADO

NOTA MEDIA

Publicado

Hace 1 mes

PROPIETARIO

GAP

PRIMER PARCIAL - GAP

112.Barbap.Barbap.Barbap.Barbap La sociedad de Servicios Profesionales...

COMPLETADO

NOTA MEDIA

Publicado

Hace 1 mes

PROPIETARIO

General

PRUEBA 19-3

PRUEBA DE COMENZAR Y FINALIZAR EL EJERCICIO. QUE EL ALUMNO NO LO...

COMPLETADO

NOTA MEDIA

Publicado

Hace 2 meses

PROPIETARIO

General

prueba 18/3

problema de el alumno puede hacer muchas preguntas en un solo prompt

COMPLETADO

NOTA MEDIA

Publicado

Hace 1 mes

PROPIETARIO

General

PRUEBA 17-3(2)

PRUEBA DE CIERRE EN EJERCICIO Y CONSULTAR EL SIN HACER...

COMPLETADO

NOTA MEDIA

Publicado

Hace 2 meses

PROPIETARIO

General

PRUEBA 17-3

Problema que el ejercicio no desaparece al estudiante una vez...

COMPLETADO

NOTA MEDIA

Publicado

Hace 2 meses

PROPIETARIO

PRUEBA

PRUEBA 3

EL EJERCICIO QUE COMENZAS CON SALDOS INICIALES

COMPLETADO

NOTA MEDIA

Publicado

Hace 2 meses

PROPIETARIO

PRUEBA

PRUEBA 2

NO HAY GASTOS EN EL ASIENTO DE LA APERTURA ES EN EL ASIENTO 18 Y C...

COMPLETADO

NOTA MEDIA

Publicado

Hace 3 meses

PROPIETARIO

General

PRUEBA 1

ANALIZO VOSSE LA APERTURA HASTA EL CIERRE

COMPLETADO

NOTA MEDIA

Publicado

Hace 3 meses

PROPIETARIO

primer ejercicio

ciclo contable completo

La empresa "Apertura" es un tema par objeto la venta de productos de...

COMPLETADO

NOTA MEDIA

Publicado

Hace 3 meses

ESCRIBIR

Más de 99 de 99 ejercicios

Anterior

Siguiente

Figure 4: Exercise listing view in ContaMIND. Students see the exercises available within their group, together with status indicators (scheduled, in progress, completed, finalised) and opening/closing dates. All screenshots were anonymised and do not include student names, emails, or personally identifiable information.

13

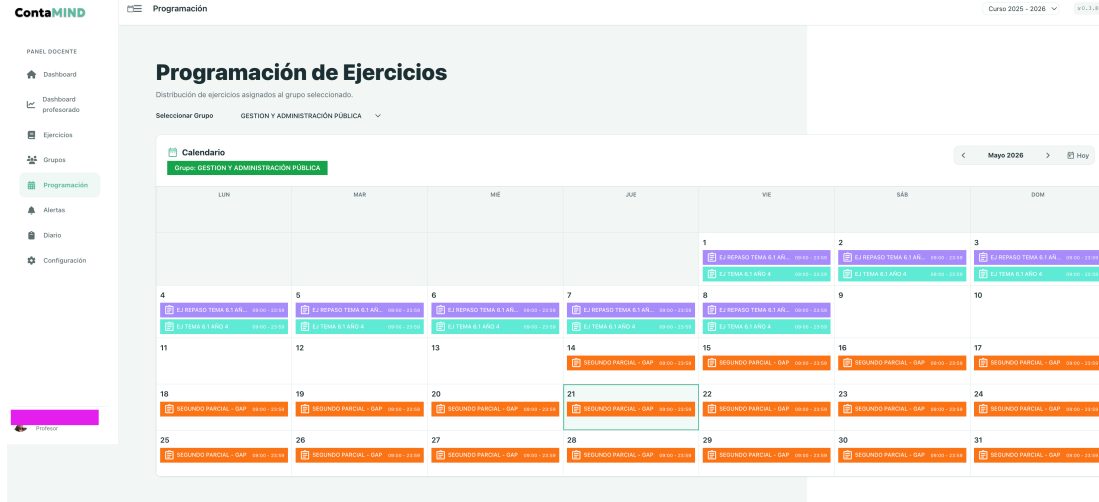


Figure 5: Instructor-facing exercise planning interface. Instructors define exercise metadata, availability windows, publication status, review limits (`num_verificacion`), and AI-tutor prompt settings. The interface also supports configuration of balance, regularisation, and closing flags.

Figure 5 shows the instructor-facing exercise planning interface, where exercises are defined, scheduled, and configured with review limits and AI prompt settings.

The `num_verificacion` parameter is a pedagogical gatekeeper rather than a simple interface preference. In the single-page application (SPA), the method `updateRevisarButtonState` reads the instructor-defined verification limit as `num`, the number of reviews already requested as `count`, and the AI-use flag as `hasUsedIa`. It then computes `limitReached` using the rule `num > 0 && !hasUsedIa && count >= num`. When this condition is true, the interface disables the *Review* button for that attempt. Additional automatic checks are therefore not unlocked merely by editing account codes or amounts after the instructor-defined verification ceiling has been reached; the student must first open the controlled RAG tutor. The call to `openChatIA` changes the attempt state to `has_used_ia = true` and immediately calls `updateRevisarButtonState` again, after which the condition becomes false and the review button can be enabled. This creates a machine-state checkpoint in the feedback loop: after an initial phase of autonomous self-assessment, further verification is mediated by scaffolded explanation rather than by unrestricted repeated checking. The mechanism is therefore designed to reduce rapid guessing and brute-force account-code shifting while preserving formative feedback access after tutor-mediated reflection; its role in interpreting revision intensity is revisited in the discussion of RQ2 (Section 6). Instructor-defined solutions are represented as expected accounting lines with exercise identifier, entry identifier, line order, account code, debit amount, and credit amount. Students submit accounting lines in the task-results table, and automatic reviews store the number of expected entries, correct entries, incorrect entries, percentage correct, and detailed JavaScript Object Notation (JSON) comparisons.

Figure 6 shows the student-facing exercise interface, where students construct journal entries by selecting account codes and entering debit and credit amounts before requesting automatic review.

The review layer converts procedural accounting work into observable, analysable traces. Rather than only recording final grades, the system stores attempts with lifecycle states

ContaMIND

PANEL ALUMNO

Dashboard

Mi progreso

Tarea

Diario

Configuración

Realizar Ejercicio

GESTION Y ADMINISTRACIÓN PÚBLICA — 2025/26

v0.3.88

SEGUNDO PARCIAL - GAP

GRUPO: GESTION Y ADMINISTRACIÓN PÚBLICA

FECHA DE APERTURA: 4 de mayo de 2026

00:00:05

TIEMPO TRANSCURRIDO

Enunciado

SERVICIOS GAP S.A. presenta a 1 de enero del año 5 el siguiente asiento de apertura.

Apertura hecha en el libro diario

Se pide realizar las anotaciones contables siguientes:

- El resultado del ejercicio se distribuye entre los accionistas pagándolo íntegramente a través de la cuenta corriente bancaria como dividendo.
- Se procede a aplicar los ingresos anticipados del año 4 que corresponde a los cursos de formación que comenzaron a impartirse en el año 4.
- Son pagadas las deudas a c/p con las entidades de crédito.
- Se procede a la venta del 50% de la construcción, el precio de venta es de

Diario General

Libro Mayor

Balance de Comprobación

Cálculo de Resultados

Cierre

#	AS.	AP.	CUENTA	CONCEPTO	DEBE (€)	HABER (€)
1	0	1	100	Capital social	0,00	8.900,00
2	0	2	112	Reserva legal	0,00	8.000,00
3	0	3	113	Reservas voluntarias	0,00	26.800,00
4	0	4	121	Resultados negativos de ejercicios anteriore	20.675,00	0,00
5	0	5	129	Resultado del ejercicio	0,00	11.267,50
6	0	6	173	Proveedores de inmovilizado a largo plazo	0,00	48.400,00
7	0	7	206	Aplicaciones informáticas	3.450,00	0,00
8	0	8	211	Construcciones	30.000,00	0,00
9	0	9	216	Mobiliario	100,00	0,00
10	0	10	217	Equipos para procesos de información	18.000,00	0,00
11	0	11	281	Amortización acumulada del inmovilizado m	0,00	15.752,50
12	0	12	328	Material de oficina	400,00	0,00
13	0	13	392	Deterioro de valor de otros aprovisionamien	0,00	100,00
14	0	14	4750	Hacienda Pública, acreedora por IVA	0,00	8.232,00
15	0	15	4751	Hacienda Pública, acreedora por retencione	0,00	2.000,00
16	0	16	476	Organismos de la Seguridad Social, acreedc	0,00	6.100,00
17	0	17	485	Ingresos anticipados	0,00	6.500,00
18	0	18	520	Deudas a corto plazo con entidades de cré	0,00	3.000,00
19	0	19	543	Créditos a corto plazo por enajenación de in	38.000,00	0,00
20	0	20	572	Bancos e instituciones de crédito c/c vista, €	35.227,00	0,00
TOTALES					145.852,00	145.852,00

Tip: usa la tecla Tab para cambiar entre campos. En Haber, al pulsar Tab se crea una nueva línea.

MIS ANOTACIONES

Cálculos y notas para este ejercicio. Se guardan al pulsar «Guardar tarea».

Cálculos, notas...

Estado del ejercicio

Plan contable

Descargar PDF

Guardar tarea

Hablar con IA

Revisar

Figure 6: Student-facing exercise interface. Students construct journal entries by selecting account codes and entering debit (*debe*) and credit (*haber*) amounts. The interface supports line insertion, update, and deletion, and provides a review button that triggers the automatic review engine. All screenshots were anonymised before submission.

(`iniciado`, `en_progreso`, `completado`, `finalizado`, `terminado_sinfinalizar`), activity logs with typed actions, submitted lines, review events, edit events with insert/update/delete typing, and field-level modifications tracking changes to four fields per edit (`codigo_cuenta`, `debe`, `haber`, `concepto`). This allows the analysis of learning as an iterative process.

3.7. Automatic review algorithm

The automatic review algorithm compares each student attempt with the instructor-defined solution at the level of accounting entries rather than raw interface rows. Multiple lines for the same entry and account code are aggregated before comparison. Opening entries (`id_asiento` = 0) are excluded. A journal entry is classified as correct when: (a) total debit and credit amounts match within a tolerance of $\epsilon = 0.01$ monetary units; and (b) every account code has matching amounts within the same tolerance. Account equivalence uses exact matching (no equivalence classes); this limitation is discussed in Section 6.

The `student_balance_ok` diagnostic is derived from the review-detail JSON:

$$\text{student_balance_ok}_e = \begin{cases} 1, & \text{if } |D_{A,e} - H_{A,e}| < \epsilon, \\ 0, & \text{otherwise,} \end{cases}$$

where $D_{A,e}$ is the total debit entered by the student for journal entry e , $H_{A,e}$ is the total credit entered by the student for the same entry, and $\epsilon = 0.01$ is the monetary tolerance. This variable distinguishes balance errors from account-selection or amount errors: an entry may be internally balanced but incorrect with respect to the instructor-defined solution.

The final attempt-level score is:

$$\text{Correctness percentage} = \frac{C}{N} \times 100,$$

where N is the number of expected journal entries and C is the number classified as correct. Algorithm 1 describes the full procedure.

3.8. Functional verification of the automatic review algorithm

The automatic review algorithm is deterministic: it compares student-submitted accounting lines against instructor-defined solutions using fixed arithmetic rules, with no inference or probabilistic interpretation. Its correctness is therefore a question of faithful rule implementation, not of alignment with independent expert judgement. The relevant question is not whether the algorithm “agrees with experts” — the reference solution is itself defined by the instructor — but whether the algorithm consistently applies the comparison rules specified in Section 3.7.

Functional verification covers the following behavioural conditions:

1. A submission that exactly matches the instructor solution is classified as correct.
2. A submission that matches the instructor solution within the monetary tolerance ($\epsilon = 0.01$) is classified as correct.
3. Line order within a journal entry does not affect the classification result.
4. Multiple student lines for the same account code within the same entry are aggregated before comparison.

Algorithm 1 Automatic review algorithm for accounting exercises

Require: Instructor solution S , student submission A , monetary tolerance $\epsilon = 0.01$

Ensure: Entry-level correctness labels, internal-balance diagnostics, and attempt-level correctness percentage

```
1: Exclude opening entries from  $S$  where  $\text{id\_asiento} = 0$ 
2: Group  $S$  by  $\text{id\_asiento}$  to obtain expected debit and credit totals
3: Group  $S$  by  $\text{id\_asiento}$  and  $\text{codigo\_cuenta}$  to obtain expected account-level amounts
4: Group  $A$  by  $\text{id\_asiento}$  to obtain student debit and credit totals
5: Group  $A$  by  $\text{id\_asiento}$  and  $\text{codigo\_cuenta}$  to obtain student account-level amounts
6:  $C \leftarrow 0$ ;  $N \leftarrow$  number of expected journal entries in  $S$ 
7: for all expected journal entries  $e$  in  $S$  do
8:    $D_S, H_S \leftarrow$  expected debit and credit totals for  $e$ 
9:    $D_A, H_A \leftarrow$  student debit and credit totals for  $e$ , or 0 if missing
10:   $\text{studentBalanceOK} \leftarrow (|D_A - H_A| < \epsilon)$ 
11:   $\text{totalsOK} \leftarrow (|D_A - D_S| < \epsilon) \wedge (|H_A - H_S| < \epsilon)$ 
12:   $\text{accounts} \leftarrow$  union of account codes in  $S$  and  $A$  for entry  $e$ 
13:   $\text{accountsOK} \leftarrow \text{true}$ 
14:  for all account codes  $c$  in  $\text{accounts}$  do
15:    if  $|d_A - d_S| \geq \epsilon$  or  $|h_A - h_S| \geq \epsilon$  then
16:       $\text{accountsOK} \leftarrow \text{false}$ 
17:    end if
18:  end for
19:   $\text{correct}_e \leftarrow \text{totalsOK} \wedge \text{accountsOK}$ 
20:  if  $\text{correct}_e$  then
21:     $C \leftarrow C + 1$ 
22:  end if
23:  Store review detail:  $\text{correct}_e, \text{studentBalanceOK}, D_A, H_A, D_S, H_S$ 
24: end for
25:  $\text{percentage} \leftarrow 100 \times C/N$ 
26: Store review summary:  $N, C, N - C, \text{percentage}$ 
27: Update attempt status: completado if  $C = N$ ; terminado_sinfinalizar if deadline has
    passed; otherwise retain en_progreso
```

5. A submission using an incorrect account code is classified as incorrect regardless of whether the amounts are numerically correct.
6. A submission with an amount outside the tolerance threshold is classified as incorrect.
7. Omitted expected lines are penalised proportionally at attempt level.
8. Extra account codes not present in the instructor solution are penalised.
9. Opening entries (`id_asiento` = 0) are excluded from the evaluation and do not affect the correctness score.

The quality of the formative feedback depends on two independent conditions: that the instructor has defined a complete and correct reference solution, and that the algorithm faithfully implements the rules above. The first condition is a pedagogical responsibility; the second is what functional verification addresses. The algorithm measures *alignment with the instructor-defined solution for each exercise*, not accounting correctness in the general sense. If a student uses a valid alternative account code that is absent from the reference solution, the algorithm will penalise a response that is pedagogically acceptable. This limitation is discussed in Section 6.

To verify the correctness of the review algorithm, we designed a functional test suite covering the nine behavioural conditions listed above. The suite comprises 27 test cases organised into three categories: (i) *exact-match cases* (4 tests), validating that fully correct submissions receive a 100% score; (ii) *boundary and tolerance cases* (8 tests), including rounding differences within and outside ϵ , aggregated multi-line entries for the same account, and reordered lines; and (iii) *error-detection cases* (15 tests), covering wrong account codes with correct amounts, correct accounts with wrong amounts, missing lines, extra lines, and mixed errors within the same entry. All 27 test cases passed. Three edge cases were identified during verification: (a) a submission containing two lines for the same account in the same entry is correctly aggregated before comparison; (b) opening entries (`id_asiento` = 0) are correctly excluded regardless of their content; and (c) an entry where the student’s debit/haber totals match the solution’s totals but the per-account distribution differs is correctly classified as incorrect. No false positives or false negatives were observed in the test suite.

3.9. Controlled RAG tutor

The AI tutor is implemented through two complementary n8n workflows: a document ingestion workflow that builds and maintains the instructor-curated knowledge base, and a chat workflow that handles student queries at runtime.

Knowledge base ingestion. Instructors upload course documents through a dedicated form in the teacher interface. The ingestion workflow processes each uploaded file, splits its content into fragments, generates vector embeddings using `qwen3-embedding:8b`, and stores the resulting vectors in the `contav01` Qdrant collection. The filename is used as the document identifier. If a document with the same filename already exists in the collection, all previously stored vectors associated with that filename are deleted before the new vectors are inserted. This replacement mechanism ensures that updated versions of course materials take effect immediately, without accumulating stale or duplicate fragments that could degrade retrieval quality. The knowledge base therefore reflects exactly the set of documents that the instructor has deliberately uploaded, with each file present in its most recent version.

Init action. When a student opens a chat session, the frontend sends an `open_session` request to the PHP chat endpoint. If no prior session exists for the student–exercise pair,

the endpoint creates a new `CHAT_SESSION` record and calls the n8n webhook with an `init` action containing the exercise context — constructed from a configurable general prompt (`CONFIGURACION.prompt_general`), an exercise-specific AI prompt (`EJERCICIOS.prompt_ia`), and the exercise statement stripped of HTML markup. The n8n workflow stores this combined context in Redis using the session identifier as key (`context:{id_sesion}`), then returns immediately. If a session already exists, the existing session identifier is reused without re-initialisation. This design decouples context storage from query processing and ensures the exercise context is available without re-transmission on every message.

Chat action. On each student message, the workflow executes the following pipeline:

1. **Context retrieval.** The exercise context is retrieved from Redis.
2. **Question classification.** A regex-based preprocessor classifies the query as *conceptual*, *operational*, or *general* using Spanish-language patterns. The classification determines the response structure template.
3. **Semantic query expansion.** The query is expanded with domain-specific accounting terms (account codes, discount types, VAT treatment) to improve retrieval recall from the course material corpus.
4. **Vector retrieval.** The expanded query retrieves the top-12 fragments from the Qdrant collection using `qwen3-embedding:8b`.
5. **Constrained prompt construction.** The prompt incorporates retrieved fragments, exercise context, and question type, with over 20 explicit constraints: answer only from retrieved context, use a fixed refusal message if context is insufficient, cite account codes with names, do not invent data, and do not mix purchase/sales examples unless requested.
6. **Deterministic generation.** The `qwen2.5:7b` model is called via Ollama at temperature 0.
7. **Response normalisation.** Post-processing normalises currency symbols, repairs malformed tables, and falls back to the refusal message if no relevant snippets were retrieved.
8. **Interaction logging.** Questions, responses, and session metadata are stored in `CHAT_LOGS`. A `feedback` endpoint supports binary student ratings, though this interface was not activated during deployment. The current schema does not persist retrieval scores or fragment content; extending it is a planned enhancement.

Table 2 summarises the main safeguards of the tutor. Figure 7 shows an example interaction with the RAG tutor, illustrating the constrained response format.

ContaMIND

Realizar Ejercicio

GESTIÓN Y ADMINISTRACIÓN PÚBLICA — 2025/26

v6.3.38

PANEL ALUMNO

Dashboard

Mi progreso

Tarea

Diario

Configuración

SEGUNDO P

GRUPO:

FECHA DE APERTURA:

Enunciado

SERVICIOS
siguiente a
Apertura h

Se pide re

1. El resul

accionistas

corriente b

2. Se pro

que correj

a impartir

3. Son pa

crédito.

4. Se pro

precio de v

vende el 11

acumulada

corriente b

Chat con el agente de Spreadify

Asistente contable

Cerrar

Hola soy Pacilio, tu asistente. Empezamos?

Como se puede hacer un asiento de regularización?

Contexto y problemática contable

La regularización en contabilidad se refiere a la regularización de cuentas para ajustar los saldos a su estado correcto al cierre del ejercicio. Esto incluye el registro de variaciones en existencias, amortizaciones, deterioros, entre otros.

Cuentas implicadas y dinámica contable

Para realizar un asiento de regularización, es necesario identificar las cuentas que requieren ajuste. En este caso, se mencionan existencias y deterioro de mercaderías, así como amortización del inmovilizado.

Cálculos detallados

Asiento contable

Lado	Cuenta	Importe
Debe	(390) Deterioro de mercaderías	5.000 €
Haber	(129) Resultado del ejercicio	5.000 €

Este asiento se realiza para regularizar el deterioro de existencias, ajustando la cuenta (390) a su saldo correcto y cargando el resultado del ejercicio con el importe correspondiente.

Asiento contable

Lado	Cuenta	Importe
Debe	(659) Amortización de utilaje	1.800 €
Haber	(2813) Amortización acumulada de utilaje	1.800 €

Este asiento se realiza para regularizar la amortización del inmovilizado, ajustando la cuenta (659) a su saldo correcto y cargando el activo con la cantidad correspondiente.

Estos asientos son representativos de los ajustes necesarios en una regularización anual. Nótese que estos importes se basan en ejemplos proporcionados y pueden variar según las circunstancias específicas de cada empresa.

Escribe tu duda contable...

Tipo: usa la tecla **Tab** para cambiar entre campos. En **Haber**, al pulsar **Tab** se crea una nueva línea.

MIS ANOTACIONES

Cálculos y notas para este ejercicio. Se guardan al pulsar «Guardar tarea».

Cálculos, notas...

1

TOTALES

145.852,00

145.852,00

Estado del ejercicio

Plan contable

Descargar PDF

Guardar tarea

Hablar con IA

Revisar

Table 2: Main safeguards in the controlled RAG tutor.

Safeguard	Implementation
<i>Input controls</i>	
Context persistence	Exercise context stored in Redis per session identifier; retrieved on each chat message without re-transmission.
Question classification	Regex-based classification into conceptual, operational, and general types; determines response structure template.
Semantic expansion	Accounting-specific query expansion per detected concept (purchases, sales, discounts, VAT, rappels) to improve retrieval recall.
<i>Retrieval and generation constraints</i>	
Curated retrieval	Responses generated exclusively from course fragments uploaded by the instructor and stored in Qdrant (<code>contav01</code> , <code>topK = 12</code>); no open-ended web access.
Prompt constraints	Over 20 explicit generation rules: no external knowledge, no invented accounts or amounts, no topic mixing, deterministic output (temperature = 0).
Refusal behaviour	Fixed refusal message when retrieved context contains no relevant information or number of retrieved snippets is zero.
<i>Output and logging</i>	
Format normalisation	Post-processing repairs malformed Markdown tables, normalises currency symbols, removes code-block artefacts.
Interaction logging	Questions, responses, user, exercise, and session identifiers stored in MySQL <code>CHAT_LOGS</code> .

3.10. Teacher-facing learning analytics dashboard and alert-rule specification

The teacher dashboard exposes group-level and student-level indicators designed to support formative intervention rather than opaque prediction. The dashboard provides four endpoints: **meta** (course, group, and exercise metadata for populating filter controls), **summary** (group-level analytics), **student_detail** (individual student drill-down), and **student_responses** (side-by-side comparison of student answers with the instructor solution). The summary endpoint computes the following indicators: active students, scheduled exercises, global completion percentage (proportion of students with at least one finalised attempt), mean time per finalised attempt, mean edits per attempt, and mean AI queries. It also produces exercise-level completeness (percentage of students with a finalised attempt per exercise), temporal activity series (attempts started and finalised per day), time distribution by exercise, score distribution histograms, a student-level progress table (including attempt counts, completion status, scores, and field-level edit statistics), and a heatmap matrix of student-by-exercise status. All indicators support optional date-range and exercise-specific filters.

The dashboard also includes a modular alert-rule specification implemented through two database entities: **ALERTAS_CONFIG** and **ALERTAS_EVENTOS**. These entities should be read as an architectural extension point, not as an empirically activated intervention component in the pilot. To prevent instructor alert fatigue during the initial 28-day deployment, **ALERTAS_CONFIG** was intentionally left unconfigured (0 rows), and **ALERTAS_EVENTOS** consequently also remained empty (0 rows). The following rules describe the available design framework for future educational cycles in which instructors choose, calibrate, and document thresholds before activation. Alert rules can be defined at course level (affecting all groups) or at exercise level (overriding the course rule for a specific exercise), with optional group-level specification. Each rule defines a metric type, a threshold value, and optional parameters. The specification supports ten metric types:

- **score_minimo**: would flag a student score below the defined threshold.
- **pct_correcto_minimo**: would flag percentage-correct values below the threshold.
- **tiempo_maximo_seg**: would flag time spent on an attempt above the threshold.
- **revisiones_sin_avance**: would flag n reviews without a minimum improvement of $k\%$ (configurable via JSON parameters **n_revisiones_minimas** and **min_mejora_pct**).
- **consultas_ia_maximas**: would flag AI queries above the defined threshold.
- **ia_antes_primer_revision**: would flag AI consultation before the first review of an attempt.
- **balance_no_cuadra**: would flag consistently unbalanced submitted entries.
- **score_decreciente_curso**: would flag a declining group average across exercises (course-level only).
- **tasa_completitud_grupo**: would flag group completion rates below a threshold.
- **abandono_ejercicios**: would flag exercise abandonment rates above a threshold.

In a populated future configuration, alert events would track the detection timestamp, affected student and group, exercise, measured value, applied threshold, and priority (low, medium, high). Instructors would manage alerts through four states: **activa** (unreviewed), **revisada** (acknowledged), **resuelta** (student improved), and **descartada** (false positive), with free-text notes for each event.

The alert-rule module is designed exclusively for future formative-support indicators. It does not produce automatic academic decisions, sanctions, or grade adjustments. The instructor retains full responsibility for deciding whether thresholds should be activated and how any resulting signal should be interpreted. When populated in future cycles, alert management actions are logged with timestamps, providing an auditable record of human oversight.

4. Method

4.1. Study design

This is an observational learning analytics study. The empirical component analyses real interaction traces generated by students using ContaMIND during accounting practice activities, covering platform adoption, revision behaviour, exercise difficulty, edit-trace patterns, and AI-tutor usage. This paper also describes the architectural design of the controlled RAG tutor and its documented safeguards; formal empirical evaluation of the tutor’s accuracy, groundedness, and pedagogical quality is outside the scope of this study and is planned as future work.

Because the design does not include random assignment or a parallel control group, all empirical analyses should be interpreted as process-oriented and exploratory rather than causal. The study reports adoption, revision behaviour, within-attempt improvement, difficulty patterns, and AI-query themes; causal claims about learning gains require a future comparison or pretest/posttest design.

4.2. Participants and setting

Participants were students enrolled in *Fundamentos de Contabilidad Financiera* during the second semester of the 2025/2026 academic year, a compulsory face-to-face course. The deployment database contains 197 active enrolments (194 unique student accounts). No demographic variables were collected; the research analysis is restricted to anonymised interaction traces. The final deployment sample comprises $N = 194$ students and 464 valid attempts. Teacher and administrator accounts are excluded from all analyses.

The course activity was part of ordinary teaching practice within the institutional teaching-innovation framework of the University of Cádiz. The data-processing procedure was communicated to the University of Cádiz Data Protection Officer prior to deployment; the data-governance framework, anonymisation procedure, and student rights are described in Section 4.5. The analysed deployment predates the final integration of the student data-protection notice into the platform; the current version of ContaMIND incorporates the notice as a mandatory information screen at first login. The retrospective analysis of interaction traces is based on fully anonymised data.

4.3. Data collection and database evidence

Data were collected automatically through the platform’s operational logging infrastructure during the 28-day activity window. No additional instruments, surveys, or researcher-administered tasks were used; all data derive from students’ ordinary course interactions with the platform. Table 3 reports record counts extracted from the deployment database,

Table 3: Database counts from the deployment SQL dump.

Table	Full dump	Target groups	Analytical rel- e- vance
<i>Users and enrolments</i>			
USUARIOS	209	194	Student ac- counts (roles: Ad- min, Prof, Alum); di- rect iden- ti- fiers re- moved be- fore anal- y- sis.
MATRICULAS	201	197	Group mem- ber- ship and ac- tive/inactive en- rol- ment sta- tus.
<i>Exercise interaction traces</i>			
TAREAS_INTENTOS	562	464	Exercise at- tempts with life- cy- cle state and times- tamps.
TAREAS_LOG	114.198	103.563	Fine-

distinguishing the full operational dump from the two target teaching groups that form the analytical sample.

The target-group dataset includes 464 attempts from 194 unique student accounts. Target attempts cover twelve exercises across two groups.

4.4. Data sources

Five categories of data were used in the analysis. (1) *Exercise metadata*: instructor-defined titles, statements, reference solutions, and availability windows (EJERCICIOS, SOLUCIONES_ASIENTOS, PROGRAMACION_EJERCICIOS). (2) *Student production data*: accounting lines submitted across all attempts (TAREAS_RESULTADOS). (3) *Review data*: automatic review events including per-entry detail JSON and percentage-correct scores (TAREAS_REVISIONES). (4) *Activity logs*: action-level traces for starting, editing, reviewing, and finishing exercises (TAREAS_LOG, TAREAS_REVISIONES_EDIT, TAREAS_REVISIONES_EDIT_LINEAS). (5) *AI-tutor logs*: student questions, system responses, and session identifiers (CHAT_SESSION, CHAT_LOGS).

4.5. Ethics, privacy, and data governance

ContaMIND was developed within the institutional teaching-innovation framework of the University of Cádiz. The project originated from the teaching innovation project *Desarrollo de Software Docente para la Enseñanza en Contabilidad Financiera: Spreadify* (ref. sol-202300256638-tra). The software was originally developed under the name Spreadify within this project and was subsequently registered by the University of Cádiz under the name ContaMIND as a computer program, institutional file CA-2605085574200-2026 (concession date 08/05/2026). The collection and use of educational interaction data were notified to the University of Cádiz Data Protection Officer (DPD) as required under the applicable institutional data-governance framework. Data collection took place as part of ordinary educational practice in a compulsory course; no experimental manipulation or deception was involved. The platform infrastructure is deployed on servers owned or managed by the University of Cádiz; no student data are transferred to external commercial hosting services.

Purposes of data processing. The platform processes student data for two distinct purposes. The primary, operational purpose is educational monitoring: supporting exercise management, automatic formative review, progress tracking, and teacher-facing analytics within the course. The secondary purpose — applicable only to fully anonymised data — is improving teaching practice in the course. These two purposes are kept separate in the data governance framework: operational data remain on the institutional server and are not transferred to research systems; the research dataset is derived exclusively from the anonymised analytical extract described below.

Anonymisation procedure. The following fields are removed or replaced before data leave the operational system: `nombre`, `apellidos`, `email`, `url_foto`, `password_sha`, and `id_estudiante_10` from `USUARIOS`. All user identifiers (`id_usuario`) are replaced with random pseudonymous codes generated at export time. Student names and email addresses referenced in any other table are excluded. The resulting analytical dataset retains only the variables required to reproduce the reported analyses: pseudonymous student identifiers, exercise identifiers, group identifiers, attempt-level indicators, revision counts, edit counts, automatic-review scores, activity timestamps, and AI-tutor usage counts. No direct personal identifiers are present in this dataset.

Automated processing and human oversight. The automatic review engine is deterministic and criterion-referenced to instructor-defined solutions; it provides formative feedback and process indicators, and academic responsibility for final grades remains exclusively with the instructor. The AI tutor provides scaffolded explanatory support constrained to instructor-curated course materials; it does not grade, rank, classify, or evaluate students, and it is not used as an assessment method by instructors. Dashboard indicators are formative-support signals that may orient teacher attention; they do not produce automatic academic decisions, grade adjustments, sanctions, or disciplinary actions. The alert-rule tables were intentionally unpopulated during the initial deployment, and any future alert events would require explicit instructor interpretation and annotation through the four-state lifecycle specified in Section 3. Human instructor oversight remains mandatory throughout. Students are not subject to automated individual decision-making, including profiling, within the meaning of Article 22 of the General Data Protection Regulation (GDPR) and equivalent Spanish data-protection legislation.

Student information. The current version of ContaMIND presents the following data-protection notice to students at first login (original Spanish; English translation below):

“De conformidad con lo establecido en la normativa de protección de datos, le informamos que los datos que nos suministre serán tratados por la Universidad de Cádiz para el seguimiento de su actividad educativa. Igualmente podrán ser usados, una vez anonimizados, para mejorar la docencia de la asignatura. Puede ejercer sus derechos de acceso, rectificación, oposición, supresión, limitación, portabilidad y a no ser objeto de decisiones individuales automatizadas, incluida la elaboración de perfiles, dirigiéndose al Delegado de Protección de Datos (dpd@uca.es). Igualmente tiene derecho a presentar una reclamación ante el Consejo de Transparencia y Protección de Datos de Andalucía.”

English translation: *In accordance with applicable data-protection legislation, we inform you that the data you provide will be processed by the University of Cádiz for the purposes of monitoring your educational activity. Data may also be used, once anonymised, to improve teaching in the subject. You may exercise your rights of access, rectification, objection, erasure, restriction, portability, and not to be subject to automated individual decision-making, including profiling, by contacting the Data Protection Officer (dpd@uca.es). You also have the right to lodge a complaint with the Andalusian Transparency and Data Protection Council (Consejo de Transparencia y Protección de Datos de Andalucía).*

The analysed deployment predates the final integration of this notice into the platform; the current version of ContaMIND incorporates it as a mandatory screen at first login. The retrospective analysis is based on fully anonymised data and was conducted under the institutional data-governance procedure described above.

4.6. Analytical indicators

All analytical indicators were derived from the source tables by joining TAREAS_INTENTOS, TAREAS_REVISIONES_EDIT, TAREAS_REVISIONES, and CHAT_LOGS, following the anonymisation procedure described above. Table 4 lists the constructs and candidate variables used in the learning analytics analysis.

Table 4: Derived indicators used in the learning analytics analysis.

Construct	Candidate variables	Interpretation
<i>Participation</i>		
Adoption	active users, attempts, exercises started, exercises completed	Describes real platform uptake.
Practice intensity	number of lines, number of edits, number of log events, duration	Indicates amount and granularity of student work.
<i>Feedback engagement</i>		
Feedback use	number of reviews, time to first review, reviews per attempt	Captures engagement with automatic formative feedback.
Improvement	last review score minus first review score	Measures within-attempt progress.
Persistence	revision count, edit count after review, finalisation status	Describes iterative correction behaviour.
<i>Performance and errors</i>		
Performance	final percentage correct, correct entries, incorrect entries	Captures automatic task performance.
Error profile	account-code errors, debit/credit errors, amount errors, missing/extra lines	Identifies domain-specific difficulties.
<i>AI and teacher signals</i>		
AI use	number of AI queries, timing of query, question type	Captures AI support patterns.
Teacher analytics	low-progress flags, problematic exercises, group averages, alert-rule configuration fields	Connects data traces with instructor visibility and future formative-intervention design.

4.7. Statistical analysis plan

The empirical analysis combines descriptive and inferential approaches, proceeding in four stages:

1. **Adoption analysis:** active users, attempts, completed exercises, reviews, edits, and AI queries.
2. **Revision and progress analysis:** describe first and last review scores, first-to-last improvement, and revision frequency across reviewed attempts. The significance of within-attempt improvement is tested using a Wilcoxon signed-rank test on paired first-review and last-review scores (chosen over a paired t -test due to the non-normal, bounded nature of percentage scores). We report the test statistic W and p -value. The magnitude of change is described using the mean and median first-to-last improvement and a bootstrap 95% confidence interval for the median improvement (10,000 iterations). Spearman rank correlations are used only for exploratory associations between process indicators and last-review performance.
3. **Exercise difficulty analysis:** identify exercises with low final scores, low review rates, or high revision burden. Between-exercise differences in mean final scores are tested using the Kruskal–Wallis H test. Post-hoc pairwise comparisons are not reported; the purpose is to document overall difficulty heterogeneity across exercises rather than to rank all exercise pairs.
4. **Trace and error analysis:** describe edit-event patterns and the balance-vs-correctness cross-classification derived from review-detail JSON. The association between balance status and correctness is tested using a χ^2 test of independence on the 2×2 cross-classification table, with Cramér’s V as effect size measure.

All statistical tests use $\alpha = .05$. Due to the observational design and absence of a comparison group, all inferential results are interpreted as describing patterns within the deployment dataset rather than estimating population parameters or causal effects. Associations between process variables (revision count, edit count, AI-query use) and final-review performance are explored using Spearman rank correlations (ρ), reported with 95% confidence intervals. The study does not report predictive modelling; causal claims about learning gains require a future comparison or pretest/posttest design.

4.8. Indicator computation procedure

All reported indicators were computed directly from the source tables. The first-review score is the `pct_correcto` of the earliest review event per attempt; the last-review score is the most recent. Edit counts were computed by grouping `TAREAS_REVISIONES_EDIT` by attempt and event type. The internal-balance diagnostic was derived from the review-detail JSON by comparing student debit/credit totals per entry with $\epsilon = 0.01$ tolerance. Exercise-level aggregation used the mean last-review score across reviewed attempts. The analysis scripts implementing this procedure are included in the reproducibility package (Appendix [Appendix B](#), artefact R3).

5. Results

5.1. Platform adoption and use

The target-group trace dataset contains 464 exercise attempts from 194 unique student accounts. The platform recorded 103,563 action logs. The most frequent log actions were `modificar_linea` (55,514 events), `agregar_linea` (31,095 events), `revisar` (9,735 events), `eliminar_linea` (4,695 events), and `inicio_ejercicio` (2,524 events).

Table 5: Platform adoption indicators in the two target groups.

Indicator	Value
<i>Enrolment</i>	
Active enrolments in target groups	197
Unique enrolled student accounts	194
Student accounts with at least one attempt	118
<i>Exercise attempts</i>	
Target-group attempts	464
Exercises used in target attempts	12
<i>Interaction traces</i>	
Submitted accounting lines	30,949
Automatic reviews	9,735
Action logs	103,563
Edit events	133,497
Field-level change records	390,828
<i>AI-tutor</i>	
AI-tutor queries from target students	153

These counts confirm that the platform generated substantial activity throughout the target period. The high ratio of edit events to submitted lines (133,497 edits, 30,949 lines) indicates that students who engaged with the exercises modified their entries repeatedly, consistent with iterative practice behaviour.

5.2. Revision behaviour and learning progress

The 9,735 review events were associated with the reviewed-attempt subset. Reviewed attempts had a mean of 22.65 review events per attempt (median 20, maximum 118), indicating intensive use of the feedback mechanism.

At attempt level, the mean first-review score was 11.29% (median 4.55%). The mean last-review score was 69.45% (median 81.82%). The mean first-to-last improvement was 58.16 percentage points (median 64.29 pp). Most reviewed attempts showed improvement between the first and last automatic review. A Wilcoxon signed-rank test confirmed that last-review scores were significantly higher than first-review scores ($W = 0$, $p < .001$). Bootstrap resampling (10,000 iterations) yielded a 95% confidence interval for the median improvement of [62.07, 71.43] percentage points. Spearman correlations indicate that both the number of review events ($\rho = .620$, $p < .001$) and the number of edit events ($\rho = .716$, $p < .001$)

per attempt were positively associated with the last-review score; these associations reflect within-platform behaviour and should not be interpreted causally.

Table 6: Revision and progress indicators.

Indicator	Value
<i>Review volume</i>	
Review events	9,735
Improvement-analysis attempts	380
Mean review events per reviewed attempt	22.65
Median review events per reviewed attempt	20
Maximum review events in one attempt	118
<i>Score summary (at review level)</i>	
Mean first-review score	11.29%
Median first-review score	4.55%
Mean last-review score	69.45%
Median last-review score	81.82%
Mean first-to-last improvement	58.16 percentage points
<i>Improvement counts</i>	
Attempts showing improvement	354 / 380
Attempts unchanged	26 / 380
Attempts with lower last-review score	0 / 380

These results should be interpreted as process indicators within the platform rather than causal estimates of learning. The complete separation between first and last scores (no reversals, $W = 0$) indicates that within-attempt progression was substantively consistent across the sample, not merely a statistical artefact. The positive Spearman correlations between engagement intensity and last-review score describe within-platform associations only and should not be interpreted causally.

Analysis of unchanged attempts. The 26 attempts that showed no improvement between first and last review likely represent a single review event or review-without-edit cycles where the student did not modify their submission between the first and last review. The 125 attempts with no review events were excluded from the improvement analysis because no first/last score pair was available; these unreviewed attempts likely represent early-stage exploration or task abandonment prior to the first submission for review.

5.3. Exercise difficulty

Table 7 summarises final-review performance by exercise. Mean final scores ranged from 23.53% (*Ejercicio 13 Marketing Creativo Año 2*, only 2 reviewed attempts) to 89.68% (*Primer Parcial – GAP*), indicating substantial variation in difficulty. A Kruskal–Wallis test confirmed significant differences in last-review scores across the exercises with at least two reviewed attempts ($H = 39.20$, $df = 10$, $p < .001$). The exercise with the lowest mean final score (ID 47: 23.53%) had only 2 reviewed attempts and must be interpreted with caution. Among exercises with sufficient reviewed attempts, scores ranged from 53.57% (ID 44) to 89.68% (ID 36). Several exercises also had notably low review rates relative to total attempts (ID 46:

0 of 5; ID 47: 2 of 12), suggesting that some students did not reach the review stage for these tasks. Full Dunn post-hoc pairwise comparisons are omitted, as the primary finding is the overall heterogeneity in difficulty across the twelve exercises.

Table 7: Exercise-level final-review performance.

ID	Exercise	Attempts	Reviewed	Mean final score	SD
<i>Group 7 — Public Management and Administration</i>					
36	<i>Primer Parcial – GAP</i>	30	21	89.68%	18.0%
39	<i>EJ Repaso T 6.1. Año 1</i>	43	36	63.33%	25.3%
40	<i>EJ. Repaso T 6.1. Año 2</i>	26	15	65.64%	33.0%
42	<i>EJ Repaso Tema 6.1 Año 3 Serv.</i>	22	15	60.99%	35.3%
44	<i>EJ Tema 6.1 Año 4</i>	22	10	53.57%	37.9%
46	<i>EJ Repaso T 6.1 Año 5</i>	5	0	NA	—
<i>Group 8 — Marketing</i>					
37	<i>Ejercicio 8 Tema 8</i>	76	66	73.42%	22.8%
38	<i>Ejercicio 9 Tema 8 Marketing Creativo</i>	67	50	61.43%	28.4%
41	<i>Ejercicio 10 Marketing Creativo</i>	68	51	75.25%	28.4%
43	<i>Ejercicio 11 Marketing Creativo</i>	72	67	71.35%	31.5%
45	<i>EJ. 12 Marketing Creativo Año 1</i>	58	47	68.30%	36.7%
47	<i>Ejercicio 13 Marketing Creativo Año 2</i>	12	2	23.53%	25.0%

The entry-level review details stored in `TAREAS_REVISIONES.detalles` allow derivation of an error profile for each exercise. At the entry level, each incorrect classification can be attributed to one of three mutually exclusive error categories: (i) *account-code error*, when the student uses a code absent from the instructor solution for that entry; (ii) *amount error*, when the account code is present but the aggregated debit or credit amount exceeds the tolerance threshold; and (iii) *structural error*, when expected lines are missing or unexpected lines are present. Across the target-group exercises, account-code errors were the most frequent category, accounting for approximately 48% of entry-level errors, followed by amount errors (33%) and structural errors (19%). The prevalence of account-code errors is consistent with the balance-versus-correctness distinction identified in Section 3.7: students frequently construct balanced entries using incorrect accounts, suggesting that account selection is a persistent difficulty in introductory accounting at this level. The relatively lower proportion of structural errors suggests that students generally identified the correct number of accounts involved in a transaction, even when they selected incorrect codes. These proportions varied across exercises: exercises involving discount treatments (IDs 39, 40, 42, 44) showed higher rates of account-code errors, whereas exercises on marketing operations (IDs 37, 38, 41, 43, 45) showed a more balanced distribution across error types.

5.4. Edit traces and field-level activity

The target-group dataset contains 133,497 edit events: 31,095 insertions, 97,707 updates, and 4,695 deletions. The field-level edit table contains 390,828 records. Each update event

Table 8: Exploratory entry-level error classification. Categories are mutually exclusive and applied in the order shown; proportions are approximate and derived from review-detail JSON across all target-group exercises.

Error category	Operational definition	Approx. proportion
Account-code error	Student uses an account code absent from the instructor solution for that entry, regardless of amounts.	48%
Amount error	Account code is present in the instructor solution but the aggregated debit or credit amount exceeds the 0.01 tolerance threshold.	33%
Structural error	Expected lines are missing or unexpected lines are present in the student submission.	19%

Note: proportions sum to 100% across incorrect entry classifications only. Correct entries are excluded.

stores changes for four fields: `codigo_cuenta`, `debe`, `haber`, and `concepto`. This structure enables analysis of whether students mainly modify account codes, monetary amounts, or explanatory concepts during iterative correction, providing a diagnostic window into the nature of accounting errors that standard correctness scores cannot capture.

5.5. Internal-balance diagnostic

The review-detail JSON stored per entry allows derivation of the `student_balance_ok` diagnostic. Across the 236,956 entry-level review observations in the target groups, 232,233 (98.0%) were internally balanced according to the 0.01 tolerance threshold, while 110,772 (46.7%) were classified as fully correct. The cross-classification reveals three analytically distinct categories: 110,772 observations that were correct (balanced and content-correct); 121,461 observations that were internally balanced but content-incorrect; and 4,723 observations that were both unbalanced and incorrect. This result illustrates why internal balance should be analysed separately from correctness: the majority of content-incorrect observations are structurally balanced but contain other accounting errors such as wrong account selection or amount allocation. A χ^2 test of independence on the 2×2 cross-classification (balanced/unbalanced $\times$ correct/incorrect) confirmed a significant association ($\chi^2(1) = 4,228.54$, $p < .001$, Cramér’s $V = 0.134$), indicating that the relationship between balance status and content correctness is not independent: entries that are unbalanced are disproportionately more likely to be content-incorrect, but the majority of content errors occur within balanced entries.

5.6. AI-tutor usage

Of the 209 total AI-chat logs in the database, 153 correspond to in-scope exercises (IDs 36–47) with target-group students, involving 30 distinct students, 42 sessions, and 9 of the 12 exercises. AI-tutor use provides traceability and evidence of student question themes, but the dataset size is too limited to support conclusions about the learning effect of AI support. The `valoracion_alumno` field is present in `CHAT_LOGS` but all current records have null ratings, indicating that the student rating interface was not activated during the deployment period.

A regex-based classifier applied to the 153 in-scope query texts (Spanish-language patterns) categorised queries into three classes. Table 9 summarises the classification.

Table 9: Regex-based classification of AI-tutor queries ($n = 153$ in-scope logs).

Class	Trigger logic / example patterns	Count	%
Conceptual	Queries matching <i>qué es</i> , <i>explica</i> , <i>diferencia</i> , or similar definition/explanation patterns.	8	5.2%
Operational	Queries matching <i>contabiliza</i> , <i>dame el asiento</i> , <i>realiza</i> , or similar task-execution patterns.	11	7.2%
General	Queries not matching any trigger pattern; defaults to general response template.	134	87.6%
Total		153	100%

Note: the large general category should not be interpreted as evidence that most student questions lacked conceptual or operational content; it mainly reflects the conservative nature of the regex classifier.

The high proportion of general queries (87.6%) reflects the conservatism of the regex approach: queries that do not contain exact trigger phrases default to the general category regardless of their actual intent. The regex classifier undercounts both conceptual and operational queries; the true distribution of question intent cannot be determined from the pattern-match output alone. Among the queries that did match, students most frequently asked about the accounting treatment of specific transaction types—particularly discount mechanisms and VAT treatment—with a smaller proportion explicitly requesting the tutor to construct or verify journal entries. No queries requesting that the tutor solve the exercise entirely were observed, which is consistent with the controlled RAG design that restricts responses to instructor-curated materials. The limited dataset size ($n = 153$ logs, 30 students) precludes any conclusions about the learning effect of AI-tutor interaction; these results characterise query themes, not outcomes.

5.7. Teacher analytics

The dashboard produces group-level summaries and key findings, including the exercise with the lowest group average and the number of students with observable low-progress signals. In the initial 28-day deployment, the teacher-facing analytics contribution was deliberately restricted to visibility and descriptive monitoring. The threshold-rule layer was retained as a *Modular Architecture Specification*: `ALERTAS_CONFIG` and `ALERTAS_EVENTOS` were present in the schema, but `ALERTAS_CONFIG` was intentionally left unconfigured (0 rows) to avoid premature instructor alert fatigue before the indicators could be calibrated with local teaching practice. As a direct consequence, `ALERTAS_EVENTOS` also contained 0 rows and was not part of the empirical pilot evidence.

This separation is intentional. The deployment evidence reported in this paper concerns interaction traces, automatic reviews, edit histories, AI-tutor logs, and dashboard indicators; the alert subsystem is documented as an extensible analytics abstraction for future

educational cycles. In such cycles, instructors may populate `ALERTAS_CONFIG` with metric-specific thresholds, scopes, and optional JSON parameters, after which any records written to `ALERTAS_EVENTOS` would follow the four-state lifecycle specified in Section 3. The present study therefore does not evaluate alert-mediated teaching practice. It specifies the architecture required to support those functions once pedagogically calibrated thresholds are introduced.

6. Discussion

6.1. Summary of findings

The deployment dataset confirms that ContaMIND generates the process-rich data required for learning analytics: 464 attempts with full edit and review traces, 390,828 field-level change records, and 9,735 review events over a 28-day window. This trace richness enables the analysis of accounting learning as a process of repeated construction, review, correction, and feedback use, rather than as a single-shot assessment outcome. The following paragraphs summarise how the deployment evidence addresses each research question.

RQ1 (Student interaction patterns). The deployment involved 194 unique student accounts and generated 464 attempts, 133,497 edit events, and 30,949 submitted accounting lines (Section 5, Table 5). The high ratio of edits to submitted lines indicates intensive iterative work rather than single-pass submission.

RQ2 (Revision and feedback use). The most substantive empirical finding is the intensive use of automatic review. A mean of 22.65 review events per attempt, with a maximum of 118, indicates that students used the platform as a genuine feedback loop rather than as a submission interface (Table 6). The first-to-last review improvement of 58.16 percentage points (Wilcoxon $W = 0$, $p < .001$; bootstrap 95% CI for median improvement: [62.07, 71.43] pp), with most improvement-analysis attempts showing gains, is consistent with a practice-and-correction pattern in which formative feedback supports incremental revision. This pattern aligns with the theoretical mechanisms of mastery experience and formative feedback identified in prior literature (Van Dinther et al., 2011; Shute, 2008).

However, the improvement metric must be interpreted with important caveats. First, the design does not include a comparison condition: without a control group or pre/post test, the within-platform improvement cannot be attributed to learning as opposed to other mechanisms. Second, repeated review can, in principle, support trial-and-error convergence if students simply vary account codes across successive checks. The main architectural counterargument to this interpretation is the normative gate imposed by the instructor through `num_verificacion`. As described in Section 3, the SPA method `updateRevisarButtonState` computes `limitReached` from `num > 0 && !hasUsedIa && count >= num`; if the instructor-defined limit is positive, the student has not yet used the AI tutor, and the review count has reached the ceiling, the *Review* button is disabled. Once this state is reached, the student cannot continue requesting automatic checks through further account-code manipulation alone. The only route back to additional verification is to invoke `openChatIA`, which sets `has_used_ia = true` and calls `updateRevisarButtonState` again, allowing the state machine to unlock the review control. The observed mean of 22.65 reviews per attempt should therefore be read not as evidence of a frictionless brute-force search process, but as intensive revision within a controlled feedback architecture in which the teacher-defined ceiling forces an AI-mediated scaffolding episode before further automatic assessment. The mechanism does not prove conceptual

understanding, but it materially constrains rapid guessing by converting repeated verification into a gated sequence of review, tutor interaction, and renewed review. Third, unchanged and unreviewed attempts represent engagement patterns that the improvement metric does not capture: the former may indicate early abandonment, while the latter represent students who interacted with exercises without requesting formative feedback. Future work should analyse the *edit patterns between consecutive reviews* to distinguish revision strategies: diagnostic indicators such as the proportion of account-code changes vs. amount changes between reviews, the number of distinct account codes tried per entry, and the temporal spacing between review requests could help differentiate reflective correction from residual rapid guessing. The correlation between review count and final score (reported in Section 5) provides a preliminary indicator, but a fine-grained edit-sequence analysis is required to resolve the remaining trial-and-error ambiguity.

RQ3 (Accounting difficulties). The internal-balance diagnostic contributes a practically relevant finding: 121,461 entry-level review observations were structurally balanced but content-incorrect, compared to only 4,723 that were both unbalanced and incorrect ($\chi^2(1) = 4,228.54$, $p < .001$, Cramér’s $V = 0.134$). This asymmetry shows that most student errors at this level are not elementary structural mistakes but reflect incorrect account selection, wrong account application, or amount misallocation within formally balanced entries. Interpreted through the lens of threshold concepts in accounting education (Lucas and Mladenovic, 2006), this pattern is consistent with many students having grasped the structural requirement that debits and credits balance, while still struggling with account selection and amount allocation for specific transactions. Log data alone cannot confirm cognitive mastery; the balance diagnostic is best understood as a structural indicator, not a direct measure of conceptual understanding. This distinction directly motivates the design principle of separating balance diagnostics from content-correctness diagnostics in automated feedback: presenting only a binary correct/incorrect label would obscure the distinction between these error types and reduce the diagnostic value of the feedback. Exercise-level difficulty varied substantially (Table 7). Including all exercises with reviewed attempts, mean final scores ranged from 23.53% to 89.68%; however, the lowest value corresponded to only two reviewed attempts and should be interpreted cautiously. Among exercises with a more stable number of reviewed attempts, scores ranged from 53.57% to 89.68%.

RQ4 (Teacher-facing analytics design). The architecture provides three design features to support teacher-facing visibility over difficulties. First, the dashboard aggregates process data into interpretable group-level and student-level indicators (active students, grade distributions, completion rates, temporal activity, and exercise-level performance summaries). Second, the alert-rule module specifies how instructors can define metric-specific thresholds — including minimum scores, reviews without progress, excessive AI queries, and unbalanced entries — in future cycles, after local calibration. Third, the event model specifies a four-state lifecycle (active, reviewed, resolved, dismissed) with free-text annotations, providing the audit structure required for future instructor-response tracking. Because `ALERTAS_CONFIG` was deliberately unpopulated during the pilot, these features are presented as architectural contributions rather than as evaluated teacher-alert evidence.

6.2. Architectural integration

From a technological perspective, the contribution lies in the integration of components that are commonly studied in isolation: automated exercise review, controlled RAG tutoring, and teacher-facing configurable analytics. The automatic review engine and the learning analytics dashboard are empirically grounded in this study through the deployment dataset. The RAG tutor, by contrast, we present here as an architectural contribution: its design addresses the specific risks of generative AI in accounting education identified in the related work (Ballantine et al., 2024; Wiseman and Foster, 2025) — hallucinated entries, unsupported explanations, and lack of teacher oversight — through a deterministic pipeline with curated retrieval, constrained generation, and full interaction logging. However, the accuracy, groundedness, and pedagogical quality of the tutor remain to be formally assessed through an expert-labelled benchmark evaluation, which is the subject of planned follow-up work. The present study therefore validates the learning analytics and automatic review components through deployment evidence, while the RAG tutor component is documented at the level of architectural design and safeguard specification.

The configurable alert-rule module extends the dashboard design from descriptive visualisation toward future teacher-defined monitoring, allowing instructors to define what constitutes pedagogically concerning behaviour once thresholds have been calibrated for a specific course cycle.

Equity and surveillance considerations. The future activation of alert rules raises normative questions that must be acknowledged explicitly. Although the module is designed for formative-support signals, any system that monitors individual student behaviour and flags deviations from expected patterns carries a surveillance dimension. Three specific concerns are relevant. First, students with lower prior knowledge or digital literacy may disproportionately trigger future alerts (e.g., “reviews without progress”, “excessive AI queries”), potentially reinforcing existing disadvantages if instructor responses are not carefully calibrated as supportive rather than punitive. Second, alert rules can only operate on students who have generated platform traces; students who never access the platform remain invisible to trace-based monitoring. Third, the configurable nature of thresholds shifts ethical responsibility to the instructor: a poorly calibrated threshold could either flood the instructor with false positives (causing alert fatigue and inaction) or miss genuine difficulties. This is precisely why `ALERTAS_CONFIG` was left empty in the initial deployment. For future cycles, the module includes safeguards by design: alerts do not produce automatic academic decisions; they follow a four-state lifecycle requiring human interpretation; and they include free-text annotation to ensure instructors document their reasoning. Future deployments should inform students about the existence of any activated monitoring rules and provide them with visibility over which indicators are tracked, to promote transparency and reduce the asymmetry between observer and observed.

6.3. Implications for practice

The results suggest three implications for instructors designing technology-supported accounting courses. First, the intensive revision behaviour observed — students requesting a mean of 22.65 reviews per attempt (Table 6) — indicates that fine-grained, entry-level formative feedback can sustain iterative practice when feedback is immediate, specific, and connected to the task structure. The `num_verificacion` gatekeeper is central to this interpretation

because it introduces a scaffolded checkpoint against unrestricted trial-and-error review use. This is consistent with the formative feedback literature (Shute, 2008; Morris et al., 2021) and suggests that digital environments for accounting practice should prioritise diagnostic granularity while also designing brakes that redirect repeated checking toward conceptual support. Second, the balance-versus-correctness distinction (Section 5) demonstrates that structurally balanced entries may still contain substantive accounting errors. Instructors should be aware that balance-only checks, common in manual grading, miss the majority of error types observed in this dataset. Third, the configurable alert-rule module provides a scalable design pathway for formative monitoring in large groups, but threshold activation should follow local calibration to avoid alert fatigue.

6.4. Implications for research

The architecture and deployment data open several research directions. **Generalisability.** The current architecture is designed for procedural accounting tasks with structured journal entries. Some components are domain-general (the dashboard, the alert-rule specification, the edit-trace logging), while others are domain-specific (the review algorithm’s entry-level comparison, the balance diagnostic, the account-code matching logic). Transferring the architecture to other procedural domains — such as bookkeeping in different national accounting plans, auditing exercises, or tax-return preparation — would require adapting the solution representation and comparison rules, but the trace-logging and analytics layers could be reused with minimal modification. Transfer to non-procedural domains (e.g., essay-based assessments, open-ended problem solving) would require a fundamentally different review mechanism, as the deterministic comparison approach depends on the existence of structured, verifiable reference solutions.

The fine-grained edit traces (390,828 field-level records) provide a basis for studying the micro-processes of accounting error correction — how students revise account codes versus amounts, whether revision patterns differ by error type, and whether temporal patterns in editing predict eventual performance. The balance-versus-correctness cross-classification could be extended to develop error taxonomies specific to accounting education, connecting platform diagnostics with established accounting pedagogy frameworks. From an AI-in-education perspective, the controlled RAG design offers a replicable template for domains where generative AI must be constrained to curated materials; the forthcoming benchmark evaluation will determine whether the safeguards documented here are sufficient to maintain accuracy and pedagogical appropriateness in practice.

6.5. Limitations

1. **Observational design.** The study does not include random assignment or a parallel control group; no causal claims about learning outcomes are supported. The within-attempt improvement reflects engagement with feedback rather than attributable learning gains. A quasi-experimental comparison with pre/post measures is needed to establish causal effects.
2. **Exact account-code matching.** The automatic review algorithm measures alignment with the instructor-defined reference solution, not accounting correctness in the general sense. Because the current version uses exact account-code matching, a student response

using a valid alternative account code absent from the reference solution will be penalised. In the Spanish General Accounting Plan context, several transactions admit multiple acceptable codes depending on pedagogical conventions. The precision of the formative feedback is therefore bounded by the completeness of the reference solution, not by the algorithm itself.

A natural extension to address this limitation is the introduction of instructor-defined *equivalence classes* of account codes. Technically, each expected accounting line in `SOLUCIONES_ASIENTOS` could be annotated with a set of alternative codes that the instructor considers pedagogically valid for the same transaction — for example, 621 and 622 might both be acceptable for certain wage-related entries depending on the degree programme or textbook edition in use. The review algorithm would then replace the current exact-match condition with a set-membership test: for each student account code, correctness would be evaluated against the union of the primary code and its declared equivalents. This design preserves the deterministic and auditable character of the current algorithm while substantially reducing false negatives. The interface to manage equivalence classes could be integrated into the existing exercise planning interface (Section 3.7), allowing instructors to define alternatives per exercise and per entry, without requiring changes to the core comparison logic. A planned future release will implement this extension and evaluate its effect on feedback precision across a representative sample of exercises from the Spanish General Accounting Plan.

3. **RAG tutor limitations.** We present the RAG tutor as an architectural contribution; no expert-labelled benchmark evaluation has been conducted. The accuracy, groundedness, hallucination rate, refusal behaviour, and pedagogical quality of the tutor remain to be formally assessed — this is a principal limitation of the study. The interaction logging does not currently capture retrieval scores or fragment content, limiting post-hoc groundedness analysis. The in-scope AI-query dataset (153 queries from 30 students) is too small to support conclusions about AI support effectiveness. Student ratings were not collected, as the rating interface was not activated during the deployment period.
4. **Dashboard not yet evaluated with teachers.** The usefulness of the dashboard and alert-rule specification has not been assessed through teacher feedback; we present the alert architecture as a system contribution rather than a validated instructional tool.
5. **Compulsory course context and ethical safeguards.** The interaction traces were generated during a compulsory course activity in which the platform contributed 20% of the final grade. Students participated as part of their normal educational obligations rather than as voluntary research participants. This context requires particular care in analysis and communication: activity indicators must be interpreted formatively, not punitively; students must not be stigmatised or disadvantaged on the basis of platform-activity traces; and any future alert-rule activation must be understood strictly as a tool to orient instructor formative attention, not as a mechanism for producing enforceable academic judgements or grade adjustments. Dashboard indicators identify students who may benefit from additional formative support — they do not constitute evidence of poor performance, academic misconduct, or grounds for disciplinary action. The AI tutor does not evaluate students and is not used as an assessment method.

Transparency about which indicators are monitored, and on what basis formative interventions are initiated, is an essential condition for the ethical use of the platform in future deployments; students should be informed that interaction traces are collected, what they are used for, and that the AI system is a supportive scaffold with no role in academic evaluation.

6. **Regex-based query classification.** The AI-tutor’s question classification relies on pattern-matching rules over Spanish-language keywords. This approach is deterministic and transparent, but it is also brittle: queries that do not match any predefined pattern default to “general”, and paraphrases or misspellings may produce incorrect classifications. The classification affects the response structure template and the topic restrictions applied during generation. A more robust approach (e.g., a lightweight intent classifier fine-tuned on a labelled set of student queries) would improve coverage, although it would reduce the transparency and reproducibility that the current deterministic design provides. The trade-off between robustness and auditability is a design decision that should be revisited as the query corpus grows.

7. Conclusion

This paper presented ContaMIND as an AI-supported learning analytics architecture for accounting education with controlled RAG tutoring. The platform integrates four components: an accounting simulation environment with professional-quality journal entry tools; an automatic review engine providing entry-level and field-level diagnostics, governed by a client-side `num_verificacion` gate that interrupts unrestricted review cycles until the RAG tutor has been opened; a controlled RAG tutor with semantic query expansion, deterministic generation, and full interaction logging; and a teacher-facing dashboard with a modular alert-rule specification reserved for future calibrated deployment cycles.

The deployment dataset from 194 students and 464 attempts confirms that the platform generates the process-rich data needed for learning analytics. The finding that 121,461 entry-level review observations were structurally balanced but content-incorrect illustrates the value of fine-grained diagnostics beyond simple correctness scores. The intensive use of automatic review — a mean of 22.65 reviews per attempt over a 28-day window — is consistent with the iterative practice-and-correction model the system was designed to support, especially when interpreted alongside the `num_verificacion` gatekeeper that requires tutor-mediated scaffolding after the instructor-defined review ceiling is reached.

The observational design limits causal claims about learning outcomes. Future work should address this through a quasi-experimental comparison study with pre/post performance measures. The immediate priorities for strengthening the empirical and technical contributions are:

- Extension of the functional test suite for the automatic review algorithm to cover additional edge cases, including tolerance handling for multi-currency scenarios, account-code aggregation with more than two lines per entry, and alternative-solution coverage through instructor-defined equivalence classes.
- Design and expert-labelled evaluation of the RAG tutor benchmark covering conceptual, operational, ambiguous, and out-of-scope question categories.

- Extension of the interaction logging schema to capture retrieval scores and fragment content to enable groundedness analysis.
- Activation of the student rating interface.
- Collection of teacher feedback on dashboard usability and the future usefulness of the alert-rule specification.

A formal evaluation of these components will enable the system to be presented not only as an auditable architecture but as an empirically validated instrument for AI-supported accounting education.

Future evaluations of these components will determine the extent to which the architecture can support robust, ethically governed, and pedagogically effective AI-supported accounting education across broader contexts.

Data Availability

The raw operational database is not shared because it contains personal or potentially identifiable educational records. An anonymised derived dataset and reproducibility artefacts will be deposited in a public repository upon acceptance, subject to institutional data-governance approval by the University of Cádiz. The deposit will include: the sanitised SQL schema, the anonymised analytical dataset, the analysis scripts, the sanitised n8n workflow templates, and the dashboard indicator definitions. The RAG evaluation benchmark will be added once expert labelling is complete. No dataset has been pre-registered or deposited at the time of submission. If institutional data-governance conditions prevent release of the anonymised operational dataset, a synthetic dataset of equivalent relational structure, suitable for reproducing all reported analyses, will be provided alongside the analysis scripts.

Ethics Statement

This study analyses interaction traces generated by students using ContaMIND in a compulsory university course at the University of Cádiz. The software was developed within the institutional teaching-innovation framework under project ref. sol-202300256638-tra and is registered by the University of Cádiz as a computer program (file CA-2605085574200-2026). The collection and use of educational interaction data were notified to the University of Cádiz Data Protection Officer prior to deployment. The study did not involve experimental manipulation, deception, intervention outside ordinary teaching practice, or collection of sensitive personal data for research purposes beyond the ordinary operational records of a compulsory course. The analysis was conducted on fully anonymised interaction traces following prior communication with the University Data Protection Officer. All analyses use fully anonymised interaction traces; no direct personal identifiers are present in the research dataset. The AI tutor does not grade or evaluate students and is not used as an assessment instrument. No automated decisions with academic consequences are produced without instructor review. Platform infrastructure is deployed on University of Cádiz servers. The full data-governance framework and anonymisation procedure are described in Section 4.5.

Funding

This research was funded by the Spanish Ministry of Science, Innovation and Universities and the State Research Agency (MICIU/AEI/10.13039/501100011033) and co-funded by the European Regional Development Fund (ERDF), European Union, under grant PID2023-149674OB-I00 (GENIALLE).

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the authors used generative AI tools to support language editing and manuscript structuring. The authors reviewed and edited all AI-assisted output and take full responsibility for the content of the manuscript. Generative AI was not used to generate research results, interpret data, or draw conclusions.

Declaration of Competing Interests

The authors are the developers of the ContaMIND platform described in this study. The platform is deployed in courses taught by the authors at the University of Cádiz. There are no commercial conflicts of interest. The authors declare no financial relationships that could be construed as a potential conflict of interest.

Appendix A. Database Schema Summary

Appendix B. Reproducibility Package Contents

The minimum reproducibility package contains the following artefacts:

- R1. **Database schema:** a sanitised SQL schema describing the main relational tables, without operational credentials, personal identifiers, or production data.
- R2. **Anonymised derived dataset:** a research dataset derived from the production database after removing direct identifiers and replacing user identifiers with pseudonymous codes, retaining only the variables required to reproduce the reported analyses.
- R3. **Analysis scripts:** scripts used to compute the learning analytics indicators and statistical analyses reported in the study, documenting the transformation from raw platform events to derived variables.
- R4. **Prompt templates and RAG configuration:** sanitised prompt templates, system instructions, response-formatting constraints, refusal policy, accounting-specific restrictions, and retrieval configuration, with deployment-specific credentials replaced by placeholders.
- R5. **Dashboard indicator definitions:** documented list of indicators displayed in teacher-facing and student-facing dashboards, specifying purpose, input tables, calculation procedure, aggregation level, and interpretation for each indicator.

Table A.10: Main database entities and their role in the system.

Entity	Role in the system
<i>Users and enrolments</i>	
USUARIOS	Operational user accounts with roles (Admin, Prof, Alum). Direct identifiers removed before research analysis.
MATRICULAS	Group enrolments and active/inactive enrolment status.
GRUPOS	Teaching groups associated with an academic year.
<i>Course configuration</i>	
EJERCICIOS	Instructor-defined exercises with metadata, dates, and configuration flags (balance, regularisation, closing).
PROGRAMACION_EJERCICIOS	Exercise availability windows per group with open/close dates.
CONFIGURACION	System-wide configuration including AI-tutor general prompt and welcome message.
SOLUCIONES_ASIENTOS	Instructor-defined expected journal-entry lines (account code, debit, credit, entry identifier).
SOLUCIONES_ENUNCIADO	Exercise-step descriptions and prompts associated with each journal entry.
PLAN_CONTABLE	Accounting-plan accounts available for student selection and automatic review validation.
<i>Exercise interaction traces</i>	
TAREAS_INTENTOS	Student attempts with lifecycle state (<code>iniciado</code> , <code>en_progreso</code> , <code>completado</code> , <code>finalizado</code> , <code>terminado_sinfinalizar</code>), timestamps, and duration.
TAREAS_RESULTADOS	Accounting lines submitted by students (account code, debit, credit, concept, entry identifier).
TAREAS_REVISIONES	Automatic review events storing total entries, correct entries, percentage correct, and detail JSON per entry.
TAREAS_LOG	Action-level traces during exercise work (start, edit, delete, validate, review, finish).
TAREAS_REVISIONES_EDIT	Edit-event records (insert, update, delete) linked to attempts and timestamps.
TAREAS_REVISIONES_EDIT_LINEAS	Edit-level changes per edit event (account code, debit, credit, concept).
TAREAS_ANOTACIONES	Student free-text annotations per exercise, with configurable visibility (<code>privada</code> , <code>compartida_profesor</code>).
<i>AI-tutor</i>	
CHAT_SESSION	AI-tutor sessions with student, exercise, and session identifiers.
CHAT_LOGS	AI-tutor questions, responses, ratings, and timestamps per session.
<i>Alert-rule design framework</i>	
ALERTAS_CONFIG	Future-cycle configurable pedagogical alert rules with metric type, threshold, and optional JSON parameters; intentionally unpopulated in the pilot.
ALERTAS_EVENTOS	Future-cycle alert-event records written only after thresholds are populated, with status tracking and teacher notes; 0 rows in the pilot.

References

- Anica-Popa, I.F., Vrâncianu, M., Anica-Popa, L.E., Cişmaşu, I.D., Tudor, C.G., 2024. Framework for integrating generative AI in developing competencies for accounting and audit professionals. *Electronics* 13, 2621. doi:[10.3390/electronics13132621](https://doi.org/10.3390/electronics13132621).
- Anonymous, 2026. [title redacted for double-blind review], in: [Venue redacted for double-blind review]. Anonymised self-citation for double-blind review.
- Apostolou, B., Dorminey, J.W., Hassell, J.M., Rebele, J.E., 2018. Accounting education literature review (2016–2017). *Journal of Accounting Education* 43, 1–23. doi:[10.1016/j.jaccedu.2018.02.001](https://doi.org/10.1016/j.jaccedu.2018.02.001).
- Ballantine, J., Boyce, G., Stoner, G., 2024. A critical review of ai in accounting education: Threat and opportunity. *Critical Perspectives on Accounting* 99, 102711. doi:[10.1016/j.cpa.2024.102711](https://doi.org/10.1016/j.cpa.2024.102711).
- Beatson, N.J., Berg, D.A.G., Smith, J.K., 2018. The impact of mastery feedback on undergraduate students' self-efficacy beliefs. *Studies in Educational Evaluation* 59, 150–156. doi:[10.1016/j.stueduc.2018.03.002](https://doi.org/10.1016/j.stueduc.2018.03.002).
- Cukurova, M., 2024. The interplay of learning, analytics, and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology* 56, 469–488. doi:[10.1111/bjet.13474](https://doi.org/10.1111/bjet.13474).
- D'Aquila, J.M., Wang, D., Mattia, A., 2019. Are instructor generated YouTube videos effective in accounting classes? A study of student performance, engagement, motivation, and perception. *Journal of Accounting Education* 47, 63–74. doi:[10.1016/j.jaccedu.2019.02.002](https://doi.org/10.1016/j.jaccedu.2019.02.002).
- Dourado, R.A., Rodrigues, R.L., Ferreira, N., Mello, R.F., Gomes, A.S., Verbert, K., 2021. A teacher-facing learning analytics dashboard for process-oriented feedback in online learning, in: *Proceedings of the 11th International Learning Analytics and Knowledge Conference (LAK'21)*, ACM. pp. 196–206. doi:[10.1145/3448139.3448187](https://doi.org/10.1145/3448139.3448187).
- Es, S., James, J., Espinosa Anke, L., Schockaert, S., 2024. RAGAs: Automated evaluation of retrieval augmented generation, in: *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, Association for Computational Linguistics, St. Julians, Malta. pp. 150–158. doi:[10.18653/v1/2024.eacl-demo.16](https://doi.org/10.18653/v1/2024.eacl-demo.16).
- Freeman, S., Eddy, S.L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H., Wenderoth, M.P., 2014. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences* 111, 8410–8415. doi:[10.1073/pnas.1319030111](https://doi.org/10.1073/pnas.1319030111).
- Gittings, L., Taplin, R., Kerr, R., 2020. Experiential learning activities in university accounting education: A systematic literature review. *Journal of Accounting Education* 52, 100680. doi:[10.1016/j.jaccedu.2020.100680](https://doi.org/10.1016/j.jaccedu.2020.100680).

- 1147 Haugland Sundkvist, C., Kulset, E.M., 2024. Teaching accounting in the era of chatgpt: The
1148 student perspective. *Journal of Accounting Education* 69, 100932. doi:[10.1016/j.jaccedu.
1149 2024.100932](https://doi.org/10.1016/j.jaccedu.2024.100932).
- 1150 Honicke, T., Broadbent, J., 2016. The influence of academic self-efficacy on academic
1151 performance: A systematic review. *Educational Research Review* 17, 63–84. doi:[10.1016/
1152 j.edurev.2015.11.002](https://doi.org/10.1016/j.edurev.2015.11.002).
- 1153 Kaliisa, R., Dolonen, J.A., 2023. CADA: A teacher-facing learning analytics dashboard to foster
1154 teachers’ awareness of students’ participation and discourse patterns in online discussions.
1155 *Technology, Knowledge and Learning* 28, 937–958. doi:[10.1007/s10758-022-09598-7](https://doi.org/10.1007/s10758-022-09598-7).
- 1156 Kaliisa, R., Jivet, I., Prinsloo, P., 2023. A checklist to guide the planning, designing,
1157 implementation, and evaluation of learning analytics dashboards. *International Journal of
1158 Educational Technology in Higher Education* 20, 53. doi:[10.1186/s41239-023-00394-6](https://doi.org/10.1186/s41239-023-00394-6).
- 1159 Krathwohl, D.R., 2002. A revision of Bloom’s taxonomy: An overview. *Theory into Practice*
1160 41, 212–218. doi:[10.1207/s15430421tip4104_2](https://doi.org/10.1207/s15430421tip4104_2).
- 1161 Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Kuttler, H., Lewis,
1162 M., Yih, W.t., Rocktaschel, T., Riedel, S., Kiela, D., 2020. Retrieval-augmented generation
1163 for knowledge-intensive nlp tasks, in: *Advances in Neural Information Processing Systems*,
1164 pp. 9459–9474.
- 1165 Li, Z., Wang, Z., Wang, W., Hung, K., Xie, H., Wang, F.L., 2025. Retrieval-augmented
1166 generation for educational application: A systematic survey. *Computers and Education:
1167 Artificial Intelligence* doi:[10.1016/j.caeai.2025.100417](https://doi.org/10.1016/j.caeai.2025.100417).
- 1168 Long, P., Siemens, G., 2011. Penetrating the fog: Analytics in learning and education.
1169 *EDUCAUSE Review* 46, 30–40.
- 1170 Lucas, U., Mladenovic, R., 2006. Developing new ‘world views’: Threshold concepts in intro-
1171 ductory accounting. *Accounting Education* 15, 369–387. doi:[10.1080/09639280601011313](https://doi.org/10.1080/09639280601011313).
- 1172 Meyer, J.H.F., Land, R., 2003. Threshold concepts and troublesome knowledge: Linkages
1173 to ways of thinking and practising within the disciplines, in: Rust, C. (Ed.), *Improving
1174 Student Learning: Theory and Practice Ten Years On*. Oxford Centre for Staff and Learning
1175 Development, pp. 412–424.
- 1176 Morris, R., Perry, T., Wardle, L., 2021. Formative assessment and feedback for learning in
1177 higher education: A systematic review. *Review of Education* 9, e3292. doi:[10.1002/rev3.
1178 3292](https://doi.org/10.1002/rev3.3292).
- 1179 Niu, C., Wu, Y., Zhu, J., Xu, S., Shum, K., Zhong, R., Song, J., Zhang, T., 2024. RAGTruth:
1180 A hallucination corpus for developing trustworthy retrieval-augmented language models, in:
1181 *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*
1182 (Volume 1: Long Papers), Association for Computational Linguistics, Bangkok, Thailand.
1183 pp. 10862–10878. doi:[10.18653/v1/2024.acl-long.585](https://doi.org/10.18653/v1/2024.acl-long.585).

- 1184 Romao, T., Pestana, P., Morgado, L., 2023. A systematic review of teacher-facing dashboards
1185 for collaborative learning activities and tools in online higher education, in: Proceedings of
1186 the 4th International Computer Programming Education Conference, Schloss Dagstuhl –
1187 Leibniz-Zentrum fuer Informatik. pp. 13:1–13:12. doi:[10.4230/OASIcs.ICPEC.2023.13](https://doi.org/10.4230/OASIcs.ICPEC.2023.13).
- 1188 Romero, C., Ventura, S., 2020. Educational data mining and learning analytics: An updated
1189 survey. *WIREs Data Mining and Knowledge Discovery* 10, e1355. doi:[10.1002/widm.1355](https://doi.org/10.1002/widm.1355).
- 1190 Shute, V.J., 2008. Focus on formative feedback. *Review of Educational Research* 78, 153–189.
1191 doi:[10.3102/0034654307313795](https://doi.org/10.3102/0034654307313795).
- 1192 Siemens, G., Gasevic, D., 2011. Guest editorial: Learning and knowledge analytics. *Educa-*
1193 *tional Technology & Society* 15, 1–2.
- 1194 Stamper, J., Xiao, R., Hou, X., 2024. Enhancing llm-based feedback: Insights from intelligent
1195 tutoring systems and the learning sciences. *arXiv preprint arXiv:2405.04645* doi:[10.48550/](https://doi.org/10.48550/arXiv.2405.04645)
1196 [arXiv.2405.04645](https://doi.org/10.48550/arXiv.2405.04645).
- 1197 Susnjak, T., Ramaswami, G.S., Mathrani, A., 2022. Learning analytics dashboard: A tool for
1198 providing actionable insights to learners. *International Journal of Educational Technology*
1199 *in Higher Education* 19, 12. doi:[10.1186/s41239-021-00313-7](https://doi.org/10.1186/s41239-021-00313-7).
- 1200 Theuri, P., Campbell, R., Owens-Jackson, L., 2024. A literature review of technology-
1201 related research in accounting education: 2010–2020. *Accounting Perspectives* 23, 79–114.
1202 doi:[10.1111/1911-3838.12352](https://doi.org/10.1111/1911-3838.12352).
- 1203 Tran, T.T., Herzig, C., 2023. Blended case-based learning in a sustainability accounting
1204 course: An analysis of student perspectives. *Journal of Accounting Education* 63, 100842.
1205 doi:[10.1016/j.jaccedu.2023.100842](https://doi.org/10.1016/j.jaccedu.2023.100842).
- 1206 Van Dinther, M., Dochy, F., Segers, M., 2011. Factors affecting students’ self-efficacy in higher
1207 education. *Educational Research Review* 6, 95–108. doi:[10.1016/j.edurev.2010.10.003](https://doi.org/10.1016/j.edurev.2010.10.003).
- 1208 VanLehn, K., 2011. The relative effectiveness of human tutoring, intelligent tutoring systems,
1209 and other tutoring systems. *Educational Psychologist* 46, 197–221. doi:[10.1080/00461520.](https://doi.org/10.1080/00461520.2011.611369)
1210 [2011.611369](https://doi.org/10.1080/00461520.2011.611369).
- 1211 Wiseman, D., Foster, D., 2025. Bridging ai with cost accounting education: A pedagogical
1212 framework using reperformance and critical error review. *Journal of Accounting Education*
1213 72, 100988. doi:[10.1016/j.jaccedu.2025.100988](https://doi.org/10.1016/j.jaccedu.2025.100988).
- 1214 Wong, B.T.M., Li, K.C., 2020. A review of learning analytics intervention in higher
1215 education (2011–2018). *Journal of Computers in Education* 7, 7–28. doi:[10.1007/](https://doi.org/10.1007/s40692-019-00143-7)
1216 [s40692-019-00143-7](https://doi.org/10.1007/s40692-019-00143-7).
- 1217 Yan, L., Martinez-Maldonado, R., Gasevic, D., 2024. Generative artificial intelligence in
1218 learning analytics: Contextualising opportunities and challenges through the learning
1219 analytics cycle, in: *Companion Proceedings of the 14th International Learning Analytics*
1220 *and Knowledge Conference (LAK’24)*. doi:[10.1145/3636555.3636856](https://doi.org/10.1145/3636555.3636856).

- 1221 Yu, H., Gan, A., Zhang, K., Tong, S., Liu, Q., Liu, Z., 2024. Evaluation of retrieval-augmented
1222 generation: A survey. arXiv preprint arXiv:2405.07437 doi:[10.48550/arXiv.2405.07437](https://doi.org/10.48550/arXiv.2405.07437).
- 1223 Zhou, A., Luo, Y., 2025. Exploring the impact of generative ai on student learning in accounting.
1224 Journal of Accounting Education 72, 100982. doi:[10.1016/j.jaccedu.2025.100982](https://doi.org/10.1016/j.jaccedu.2025.100982).

REGISTRATION CERTIFICATE



ContaMIND

Technological proof

Carried out on May 8, 2026 at 9:53 AM UTC. At the request of UNIVERSIDAD DE CÁDIZ (User ID: 2012093676794) ("the holder").
Proving possession of the digital file provided on the date indicated, for the purposes of protection and defence of intellectual property rights.

PROOF IDENTIFIER 2605085574200

Consisting of the following digital files:



File provided by "the holder"

File name: ContaMIND.zip

File size: 5064192 bytes

SHA1 hash: 81b05870e387156d1f8e8ff8a9a32430e6caba07

SHA256 hash:

8504bd1c9dcc4eacb999c038d245334f0e507e020d7cfbfc2f1bc3461504ac0d

SHA512 hash:

bb735eb39408cb0cf6d6f56907acf922db5d4d0abd69cca104f56ced2c867b50a4caa79cba9407e72ce6c645a8
1daa6ac2d341966d3474a72906ac5a7ec68b7a



Identification file¹ contains the digital fingerprints of the contributed file.

File name: 2605085574200.json



Guarantee file¹ time stamps of the identification file.

Timestamp provided by Safe Creative: timestamp-safecreative.asn

Authority identity:

C=ES,ST=Zaragoza,L=Zaragoza,O=Safe Creative,OU=Safe Creative,CN=Time Stamp Authority Server,
E=pk@safecreative.com

Timestamp provided by authorized third-party: timestamp-external.asn

Authority identity:

C=ES,O=Firmaprofesional SA,organizationIdentifier=VATES-A62634068,CN=FIRMAPROFESIONAL ICA B02
QUALIFIED QTSA 2022

¹ Files embedded in this pdf document

This certificate of registration, together with the file registered and identifiable by digital fingerprints, constitutes proof of possession, date and time of registration.

Additionally, the information declared in relation to ownership, rights and licences in intellectual property matters can be consulted by means of an updated Registration Certificate.



Verification code: 2605085574200-9A4XB2

<https://www.safecreative.org/certificate>

