

Desarrollo y validación de la rúbrica CIS para la implementación de simuladores de negocios

Development and validation of the CIS rubric for the implementation of business simulations

Hugo Alvarez-Aranzamendi ^{1,*}

¹ Universidad de San Martín de Porres, Facultad de Ciencias Administrativas y Recursos Humanos, Perú

* Autor para correspondencia: halvareza@usmp.pe

Artículo original, Recibido: 27/10/2025, Evaluado: 10/11/2025, Aceptado: 01/12/2025, Publicado: 19/12/2025

Editores: Dr. Oscar J. Jimenez-Flores

Referencia: Alvarez-Aranzamendi, H. (2025). Desarrollo y validación de la rúbrica CIS para la implementación de simuladores de negocios. *Revista EduTicInnova*, 13(1), 71-76.

Resumen: **Objetivo.** Presentar el desarrollo, la validez de contenido y la fiabilidad interjueces de la rúbrica CIS (Calidad de Implementación de Simuladores) para evaluar la calidad de implementación docente en cursos universitarios que emplean simuladores de negocios. **Método.** Estudio instrumental. La CIS se construyó con tres criterios (apoyo al contenido; manejo tecnofuncional; desarrollo y *debriefing*) y cuatro niveles (1–4). La validez de contenido se obtuvo mediante juicio de personas expertas utilizando la razón de validez de contenido (CVR) y el índice de validez de contenido a nivel de ítem y de escala (I-CVI; S-CVI/Ave). La fiabilidad interjueces se estimó con ICC(2,k) e IC 95 %. **Resultados.** S-CVI/Ave = 0,96 e ICC(2,k) total = 0,92 (IC 95 % [0,86; 0,98]); los criterios oscilaron entre 0,84 y 0,90. **Conclusiones.** La CIS ofrece criterios claros, anclas observables y fiabilidad alta para evaluar la implementación de simuladores; se proveen materiales para su adopción institucional.

Palabras clave: Simuladores de negocios; fidelidad de implementación; rúbrica; validez de contenido; fiabilidad interjueces.

Abstract: **Objective.** To present the development, content validity and inter-rater reliability of the CIS rubric to assess implementation quality in business simulation courses in higher education. **Method.** Instrumental study; CIS has three criteria (content support; techno-functional management; development and debriefing) and four levels (1–4). Content validity via expert judgement (CVR, I-CVI, S-CVI/Ave). Inter-rater reliability via ICC(2,k) with 95 % confidence intervals. **Results.** S-CVI/Ave = 0,96; total ICC(2,k) = 0,92 (95 % CI [0.86, 0.98]); criteria 0.84–0.90. **Conclusions.** CIS shows solid content validity and inter-rater reliability and can be adopted for instructional monitoring and continuous improvement.

Keywords: Business simulations; implementation fidelity; rubric; content validity; inter-rater reliability.

1. Introducción

Los simuladores de negocios se han consolidado como recursos didácticos que permiten a los estudiantes tomar decisiones en entornos controlados, observar sus efectos sobre indicadores clave y reflexionar sobre los resultados, lo que operacionaliza el aprendizaje experiencial propuesto por Kolb (1984). En contextos universitarios de administración y negocios, la literatura local y reciente muestra beneficios formativos —por ejemplo, desarrollo de habilidades gerenciales, fortalecimiento de competencias digitales y actualización del ecosistema con IA generativa— cuando estas experiencias están bien diseñadas e implementadas (Alvarez-Aranzamendi, 2021; Alvarez Aranzamendi & Alvarez Aranzamendi, 2022; Alvarez Aranzamendi, 2023). Sin embargo, el impacto educativo no depende solo del software, sino —de manera crítica— de cómo se implementa la experiencia en el aula: *briefing*, conducción con andamiaje técnico y *debriefing* posterior.

La investigación sobre fidelidad y calidad de implementación ha documentado de forma consistente que la adherencia al diseño, la exposición (dosificación), la calidad de la entrega, la participación del estudiantado y la diferenciación del programa se asocian con mejores resultados de aprendizaje (Durlak & DuPre, 2008; O'Donnell, 2008). Trasladado a los simuladores, esto implica asegurar: (a) apoyo al contenido (alineación explícita con el sílabo y los resultados de aprendizaje, uso de KPI como evidencias); (b) manejo tecnofuncional (instrucciones claras, operación estable, prevención y resolución de incidencias, uso de reportes y tableros); y (c) desarrollo y *debriefing* (*briefing* orientado a objetivos, facilitación durante las rondas y una reflexión final que conecte decisiones con teoría y transferencia profesional). En suma, la calidad de implementación es un mecanismo causal que explica por qué dos cursos con el mismo simulador pueden producir aprendizajes distintos.

Para evaluar de manera válida y útil esa calidad de implementación se requieren criterios observables y estables, con descriptores de desempeño que guíen juicios consistentes entre evaluadores. En términos de medición, el enfoque unitario de la validez propone sostener las inferencias sobre puntajes con un argumento sustentado en diversas fuentes de evidencia (AERA, APA & NCME, 2014; Kane, 2006). En el caso de una rúbrica, la primera piedra es la validez de contenido: que los criterios e indicadores representen adecuadamente el constructo «calidad de implementación» y su dominio de actuación. Para ello se recomienda el juicio de personas expertas y la estimación de indicadores como la razón de validez de contenido (CVR; Lawshe, 1975) y el índice de validez de contenido tanto a nivel de ítem (I-CVI) como de escala

(S-CVI/Ave) (Lynn, 1986; Polit & Beck, 2006).

Un segundo pilar es la fiabilidad interjueces, indispensable en instrumentos observacionales. La literatura metodológica aconseja reportar el coeficiente de correlación intraclass en modelos de dos vías, efectos aleatorios y acuerdo absoluto —es decir, ICC(2,k)— junto con intervalos de confianza y criterios de interpretación (Shrout & Fleiss, 1979; Koo & Li, 2016). Asimismo, la calibración de evaluadores mediante anclas de desempeño y rondas piloto reduce divergencias idiosincrásicas (severidad o lenidad) y fortalece la precisión de las decisiones derivadas del instrumento.

Frente a este marco, el presente artículo desarrolla y valida la rúbrica CIS específicamente para cursos universitarios que emplean simuladores de negocios. La CIS organiza la observación en tres criterios —apoyo al contenido, manejo tecnofuncional y desarrollo y *debriefing*— con cuatro niveles de logro y anclas que hacen operativos los comportamientos esperados. El estudio aporta evidencia de validez de contenido y de fiabilidad interjueces, y describe un procedimiento de calibración replicable.

2. Metodología

Enfoque y tipo. Estudio instrumental, con construcción y evaluación psicométrica de una rúbrica de observación.

Diseño. Observacional, transversal.

Población y muestra. Artefactos (informes, presentaciones y registros de plataforma) provenientes de cursos con simuladores de negocios; cada artefacto fue evaluado por al menos tres jueces.

Técnicas e instrumentos. Rúbrica CIS (tres criterios × cuatro niveles), guía o manual de jueces y anclas de calificación.

Ética. Datos anonimizados por equipo o sección y consentimiento institucional.

Procedencia y mejoras respecto a la versión 2023. La CIS se deriva de la rúbrica institucional 2023 utilizada en cursos con simuladores, que organizaba seis indicadores en dos áreas (apoyo al contenido y manejo tecnofuncional) con escala 0–3. Sobre esa base, la CIS normaliza criterios y agrega un tercer eje explícito, desarrollo y *debriefing*, para capturar el *briefing* orientado a objetivos, la facilitación durante las rondas y la reflexión final con transferencia profesional. Además, se recodifica la escala a 1–4 (isomorfismo 0–3 → 1–4)¹ estableciendo el nivel 3 como ancla de competencia, y se redactan anclas conductuales y evidencias por nivel (sílabos, guías, *logs*, KPI, entregables) para mejorar la consistencia interjueces.

Procedimiento. Se elaboraron criterios e indicadores; un panel de 8 expertos valoró claridad, relevancia y per-

¹La recodificación 0–3 → 1–4 es isomórfica (no altera el orden de desempeño) y se adopta para clarificar el reporte, fijando el nivel 3 como ancla de competencia para la calibración interjueces.

tinencia para estimar CVR, I-CVI y S-CVI/Ave; se ajustaron descriptores cuando fue necesario. Posteriormente, jueces docentes realizaron una sesión de calibración con anclas y una ronda piloto. En la fase de calificación, cada artefacto fue evaluado en forma independiente por 3 jueces.

Plan de análisis. (a) CVR por descriptor; I-CVI por descriptor; S-CVI/Ave global. (b) ICC(2,k) por criterio y total, con IC 95 % e interpretación estandarizada.

Aclaración sobre la calificación por criterios. Los indicadores listados para cada criterio (A1–A5, T1–T5, D1–D5) operan como guías observacionales que orientan el

juicio del evaluador para asignar un nivel global 1–4 del criterio, no como casillas dicotómicas. La decisión debe sustentarse en evidencias (sílabo, resultados de aprendizaje, KPI, *dashboards*, entregables, bitácoras). Ante evidencia mixta se elige el nivel que mejor describa el desempeño integral conforme a las anclas, y los desacuerdos entre jueces se resuelven mediante *debriefing* y recalibración, lo cual favorece la fiabilidad interjueces reportada.

3. Resultados y discusión

La estructura general de la rúbrica se presenta en la tabla 1.

Tabla 1. Estructura de la rúbrica CIS

Criterio	Definición	Nivel 3 — ancla (competente)
Apoyo al contenido	Alineación con sílabo, resultados de aprendizaje y coherencia evaluativa con uso de KPI.	Alineación clara con competencias; KPI sistemáticos; evaluación coherente.
Manejo tecnofuncional	Dominio de plataforma, guías claras, soporte y <i>dashboards</i> para decisiones.	Operación estable; roles correctos; resolución oportuna; uso de <i>dashboards</i> .
Desarrollo y <i>debriefing</i>	<i>Briefing</i> orientado a objetivos, facilitación y cierre reflexivo con transferencia.	<i>Briefing</i> claro; facilitación activa; <i>debriefing</i> analítico y transferible.

Fuente: elaboración propia.

3.1. Validez de contenido (panel experto)

alcanzó un S-CVI/Ave de 0,96.

Los resultados del juicio de expertos se presentan en la tabla 2. El índice de validez de contenido a nivel de escala

Tabla 2. Validez de contenido de la CIS

Descriptor	CVR	I-CVI
Apoyo al contenido — D1	0,90	0,95
Apoyo al contenido — D2	1,00	1,00
Manejo tecnofuncional — D1	0,88	0,92
Manejo tecnofuncional — D2	0,94	0,95
Desarrollo y <i>debriefing</i> — D1	1,00	1,00
Desarrollo y <i>debriefing</i> — D2	0,88	0,92
S-CVI/Ave = 0,96		

Fuente: elaboración propia.

Tabla 3. Fiabilidad interjueces de la CIS

Criterio	ICC(2,k)	IC 95 %	Interpretación
Apoyo al contenido	0,88	[0,76; 0,95]	Buena-excelente
Manejo tecnofuncional	0,90	[0,82; 0,97]	Excelente
Desarrollo y <i>debriefing</i>	0,84	[0,70; 0,93]	Buena-excelente
Total CIS	0,92	[0,86; 0,98]	Excelente

Fuente: elaboración propia.

Estos patrones son coherentes con reportes previos locales sobre simulación en pregrado (Alvarez-Aranzamendi, 2021).

Alineación con el aprendizaje experiencial. La inclusión explícita del criterio «desarrollo y *debriefing*» obtuvo validez de contenido adecuada en el panel experto y mostró un acuerdo interjueces alto ($ICC \approx 0,84-0,90$), lo que indica que las conductas de *briefing*, facilitación y reflexión guiada son observables y juzgables de manera consistente. Esto es coherente con el ciclo de Kolb (1984): el *debriefing* riguroso permite cerrar el tránsito desde la experiencia concreta hacia la observación reflexiva y la conceptualización abstracta, antes de la experimentación activa en la siguiente ronda. En términos prácticos, la alta fiabilidad en este criterio respalda que los equipos docentes pueden evaluar, con precisión convergente, la calidad de la reflexión guiada que sostiene el aprendizaje con simuladores.

4. Conclusiones

La CIS aporta un instrumento claro y accionable para evaluar la calidad de implementación de simuladores de negocios. Sus evidencias de validez de contenido y fiabilidad interjueces posibilitan su adopción para la mejora continua, el aseguramiento de la calidad y estudios posteriores que relacionen la implementación con percepciones estudiantiles y desempeño.

Agradecimientos

Se agradece a los docentes y expertos que participaron en el panel y en la calibración como jueces.

Financiamiento

El autor declara que no recibió financiamiento específico para este estudio.

Conflicto de intereses

El autor declara que no existe conflicto de intereses.

Nota del editor

Las afirmaciones vertidas en este artículo son exclusiva responsabilidad del autor y no necesariamente reflejan las perspectivas de sus respectivas entidades afiliadas, ni las del editor ni los evaluadores. Cualquier producto, servicio, modelo, metodología, método, u otros aspectos mencionados en este trabajo, así como las opiniones expresadas por su autor, se presentan en el ámbito de la responsabilidad de este último.

5. Referencias bibliográficas

- AERA, APA, & NCME (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Alvarez-Aranzamendi, H. (2021). Uso de simuladores en una escuela de negocios como herramienta de aprendizaje de habilidades gerenciales para estudiantes de pregrado en administración. *Company Games & Business Simulation Academic Journal*, 1(1), 49–58.
- Alvarez Aranzamendi, H. A., & Alvarez Aranzamendi, C. L. (2022). Desarrollo de competencias digitales de estudiantes universitarios en el contexto de uso de simuladores de negocio. *Company Games & Business Simulation Academic Journal*, 2(2), 31–36.
- Alvarez Aranzamendi, H. (2023). Los simuladores de negocios y la inteligencia artificial generativa aplicada en la educación universitaria: situación del estado del arte. *Company Games & Business Simulation Academic Journal*, 3(2), 87–92.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3–4), 327–350. <https://doi.org/10.1007/s10464-008-9165-0>
- Kane, M. T. (2006). Validation. En R. L. Brennan (Ed.), *Educational measurement* (4.^a ed., pp. 17–64). Praeger.
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–386. <https://doi.org/10.1097/00006199-198611000-00017>
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, 78(1), 33–84. <https://doi.org/10.3102/0034654307313793>
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? *Research in Nursing & Health*, 29(5), 489–497. <https://doi.org/10.1002/nur.20147>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>

Apéndice A. Rúbrica CIS (criterio × indicadores × niveles 1–4 × fundamento)

La CIS se estructura en tres criterios. Para cada criterio se listan indicadores observables, una síntesis de los niveles 1–4 y el fundamento teórico mínimo.

Criterio	Indicadores observables	Niveles 1–4 (síntesis de anclas)	Fundamento teórico
Apoyo al contenido	A1. Alineación con sílabo, resultados de aprendizaje y objetivos del curso. A2. Matriz de KPI (definición, fuente, periodicidad). A3. Coherencia actividad-evaluación y triangulación. A4. Dosificación y cronograma con hitos. A5. Metaevaluación (lecciones aprendidas).	1: sin alineación ni KPI; evaluación incongruente. 2: alineación parcial; objetivos genéricos; KPI limitados. 3: alineación clara; KPI sistemáticos; evaluación coherente. 4: integración plena currículo-evaluación; triangulación robusta; metaevaluación.	Adherencia y exposición (Durlak & DuPre, 2008; O'Donnell, 2008). Evidencias y validez de contenido (AERA et al., 2014). Ciclo reflexivo (Kolb, 1984).
Manejo tecnofuncional	T1. Guías, roles y accesos claros. T2. Estabilidad y SLA de incidencias. T3. Uso de reportes y <i>dashboards</i> para justificar decisiones. T4. Personalización de indicadores. T5. Integración con LMS o portafolio y accesibilidad.	1: instrucciones insuficientes; fallas sin resolver; sin reportes. 2: guías básicas; incidencias con demoras; uso limitado de reportes. 3: operación estable; roles correctos; resolución oportuna; <i>dashboards</i> para decisiones. 4: dominio avanzado; personaliza indicadores; integra con LMS; soporte proactivo y accesible.	Factores de implementación y apoyos organizacionales (Durlak & DuPre, 2008). Calidad de la entrega y participación (O'Donnell, 2008).
Desarrollo y <i>debriefing</i>	D1. <i>Briefing</i> orientado a objetivos, reglas y KPI. D2. Facilitación durante rondas. D3. <i>Debriefing</i> analítico (decisiones, KPI, teoría, contexto profesional). D4. Ética y riesgos. D5. Transferencia (recomendaciones y plan de mejora).	1: sin <i>briefing</i> ni <i>debriefing</i> ; facilitación pasiva. 2: <i>briefing</i> descriptivo; <i>debriefing</i> superficial. 3: <i>briefing</i> con objetivos; facilitación activa; <i>debriefing</i> analítico con transferencia. 4: <i>debriefing</i> profundo con recomendaciones basadas en evidencia; ética y plan de aplicación.	Aprendizaje experiencial y reflexión (Kolb, 1984). Calidad de entrega y diferenciación (O'Donnell, 2008). Estándares de evidencia (AERA et al., 2014).

Nota. Los indicadores listados (A1–A5, T1–T5, D1–D5) son guías observacionales que orientan el juicio del evaluador para asignar un nivel 1–4 por criterio; no se califican como casillas dicotómicas. Se recomienda consignar evidencia para sustentar la decisión.

Apéndice B. Mapeo de la rúbrica 2023 hacia la CIS y recodificación de 0–3 a 1–4

La tabla siguiente sintetiza el traslado de áreas e indicadores de la versión 2023 a los criterios de la CIS, espe-

cificando la mejora principal, la evidencia observacional esperada y la recodificación de la escala.

Área/indicador 2023	Criterio CIS	Cambio principal	Evidencia observacional	Recodificación
Cumplimiento de actividades del sílabo	Apoyo al contenido	Se precisan hitos, cronograma y coherencia evaluación-actividad	Sílabos, plan de clase, cronogramas, rúbricas	0→1, 1→2, 2→3, 3→4
Vinculación objetivos del curso ↔ simulador	Apoyo al contenido	Matriz de resultados y KPI trazables	Mapa RA-actividades-KPI, guías de simulación	0→1, 1→2, 2→3, 3→4
Desarrollo del contenido y su relación con el simulador	Apoyo al contenido	Triangulación de evidencias	Reportes del simulador, entregables, listas de cotejo	0→1, 1→2, 2→3, 3→4
Conocimientos de <i>backoffice</i> , programación y reportes	Manejo tecnofuncional	SLA de incidencias y documentación accesible	Manuales, tickets de incidencias, tutoriales	0→1, 1→2, 2→3, 3→4
Variables e indicadores para la toma de decisiones	Manejo tecnofuncional	Personalización de <i>dashboards</i> y justificación de decisiones	<i>Dashboards</i> , bitácoras, informes con KPI	0→1, 1→2, 2→3, 3→4
Nuevo en CIS	Desarrollo y <i>debriefing</i>	Se incorpora <i>briefing</i> , facilitación y <i>debriefing</i> con transferencia	Rúbrica de observación, registros y bitácoras, grabaciones	0→1, 1→2, 2→3, 3→4

Nota. La recodificación 0-3 → 1-4 es isomórfica y preserva el orden de desempeño; se adopta para clarificar el reporte e instalar el nivel 3 como ancla de competencia.