

Unidades críticas y error severo en Matemáticas Básicas de Ingeniería: análisis por reactivo y brecha percepción–desempeño

Critical units and severe error in Basic Engineering Mathematics: item-level analysis and the perception–performance gap



 **Dr. Antonio Rosado Capetillo**



 **Dr. Uriel Gabriel Zapata Rodríguez**



 **Dr. Israel Ortíz Vargas**

1. Facultad de Ingeniería Mecánica y Ciencias Navales, Universidad Veracruzana, México.
2. Facultad de Enfermería, Universidad Veracruzana, México.

Fecha de recepción: 15 de abril de 2026

Fecha de aceptación: 02 de junio de 2026

Fecha de publicación: 22 de julio de 2026

Resumen

Este artículo reporta un diagnóstico cuantitativo, a nivel de reactivo, del desempeño y los patrones de error en la asignatura Matemáticas Básicas de Ingeniería impartida a estudiantes de Ingeniería Mecánica Eléctrica (IME). Se analizaron registros reales de tres exámenes parciales (agosto 2025–enero 2026) correspondientes a cuatro grupos, transformados a una matriz en formato largo (una fila por reactivo) que integra unidad temática, tema, puntaje porcentual por reactivo (ScorePct) y tipo de error codificado (0: sin error; 1: procedimiento; 2: resultado; 3: fallo total; 4: parcial/incompleto). Adicionalmente, se aplicó una encuesta de cierre (N=154) con dominios

de percepción, retroalimentación, evaluación, satisfacción, hábitos y dificultades. Los resultados evidencian alta heterogeneidad por unidad: Trigonometría Básica y Operaciones presenta el menor ScorePct medio (24.43%) y la mayor tasa de error severo (91.18%), mientras que Números Reales exhibe un desempeño alto (81.99%) y baja severidad (19.12%). La encuesta muestra percepciones relativamente altas por unidad (≈ 3.75 – 3.84 en escala 1–5), lo que sugiere una brecha percepción–ejecución en unidades críticas (Zimmerman, 2002). Se discuten implicaciones didácticas para priorizar intervenciones formativas con retroalimentación inmediata, verificación por pasos y práctica distribuida, y se proponen líneas futuras para validación pre–post y replicación por cohortes.



Esta obra está bajo la licencia Creative Commons Attribution-NonCommercial 4.0 International.

Palabras clave: evaluación por reactivo, teoría del error, matemáticas en ingeniería, trigonometría, retroalimentación, triangulación

Abstract

This paper reports an item-level quantitative diagnosis of performance and error patterns in an introductory engineering mathematics course (Mathematics Fundamentals) taught to Mechanical–Electrical Engineering students. Three midterm exams (Aug 2025–Jan 2026) from four class groups were converted into a long-format dataset (one row per item) including topic/unit, item score percentage (ScorePct), and a coded error type (0: correct; 1: procedure; 2: result; 3: total failure; 4: partial/incomplete). A post-course survey (N=154) measured perceived mastery, feedback quality, assessment alignment, satisfaction, study habits, and perceived difficulties. Results show strong heterogeneity across units: Basic Trigonometry exhibits the lowest mean ScorePct (24.43%) and the highest severe-error rate (91.18%), whereas Real Numbers shows high mean performance (81.99%) and low severity (19.12%). Survey responses indicate relatively high perceived mastery across units (≈ 3.75 – 3.84 on a 1–5 Likert scale), suggesting a perception–execution gap in critical units. Implications for teaching include prioritizing formative micro-assessments, immediate feedback, stepwise verification, and distributed practice; future work is proposed via pre–post designs and multi-cohort replication.

Keywords: item-level assessment, error analysis, engineering mathematics, trigonometry, feedback, triangulation

Introducción

En la formación de ingeniería, los cursos matemáticos de los primeros semestres suelen operar como “materias habilitantes” y, en algunos casos, como cuellos de botella que condicionan el avance en asignaturas posteriores (circuitos, máquinas eléctricas, señales, control, métodos numéricos). En este entorno, la evaluación tradicional basada únicamente en la calificación global aporta información limitada: indica cuánto se obtuvo, pero no explica cómo se falló, qué patrones se repiten y qué contenidos concentran mayor severidad (Black & Wiliam, 1998). La teoría del error en matemáticas plantea que el error no debe interpretarse como accidente, sino como evidencia diagnóstica que puede orientar intervenciones específicas. Debido a esto, el análisis a nivel de reactivo —vinculando contenido, puntaje y tipología de error— permite transformar la evaluación sumativa en una fuente de información para mejora docente (Movshovitz-Hadar et al., 1987). Este estudio se desarrolló en el marco de la asignatura Matemáticas Básicas de Ingeniería, en la carrera de Ingeniería Mecánica Eléctrica, con cuatro grupos y tres parciales durante el periodo agosto 2025–enero 2026. La estructura real de evaluación del curso

es relevante para interpretar resultados: cada parcial evalúa conjuntos de unidades distintas, lo que impide comparar parciales como mediciones repetidas del mismo contenido; a pesar de esto, sí permite caracterizar cómo cambia la naturaleza del error cuando cambia el tipo de contenido evaluado. Adicionalmente, se aplicó una encuesta de cierre para captar percepción de aprendizaje, retroalimentación, hábitos y dificultades, con el propósito de triangular evidencia objetiva (exámenes) y subjetiva (percepción) (Prince & Felder, 2006).

El artículo tuvo dos objetivos centrales:

- (i) Identificar unidades y temas críticos mediante métricas de desempeño y severidad del error, y
- (ii) Evaluar la existencia de brechas percepción–desempeño que puedan estar limitando la calibración metacognitiva del estudiante. Los hallazgos permiten priorizar intervenciones didácticas de alto impacto y plantear un modelo replicable de diagnóstico semestral basado en una matriz por reactivo.

En años recientes, la analítica educativa ha reforzado el valor de los diagnósticos finos, a nivel de ítem o reactivo, como apoyo para la toma de decisiones didácticas en contextos STEM y educación superior, particularmente cuando se busca traducir registros de evaluación en intervenciones de mejora focalizada.

Metodología

Diseño y datos

Se empleó un diseño cuantitativo no experimental (ex post facto) con análisis longitudinal intra-semestre basado en datos existentes. Se analizaron tres exámenes parciales correspondientes a cuatro grupos, con registros por estudiante y por reactivo. Los exámenes se transformaron a una matriz en formato largo (una fila por reactivo), lo que permitió operar con la unidad mínima de evidencia (reactivo) para medir desempeño y tipología de error (Schoenfeld, 1985).

Variables e instrumentación

Para cada registro alumno–parcial–reactivo se integraron: (a) unidad temática y tema del reactivo; (b) puntaje porcentual por reactivo (ScorePct); (c) tipo de error codificado: 0 (sin error/correcto), 1 (falló en procedimiento), 2 (falló en resultado), 3 (fallo total: procedimiento y resultado), 4 (parcial/incompleto). Con fines de comparabilidad, se definió “error severo” como **ErrorCode $\in \{3,4\}$** . El desempeño global por examen se operó como el porcentaje obtenido en el resultado global del examen. La categoría de “error severo” se definió a partir de la agrupación de los códigos 3 y 4, al corresponder a situaciones en las que la validez de la respuesta quedó comprometida de manera sustantiva: ya fuera por un fallo total del desarrollo o por una resolución incompleta que impidió arribar a un resultado matemáticamente interpretable.

Esta decisión se sustentó en la lógica de la teoría del error, según la cual no todos los fallos tienen el mismo peso diagnóstico, y algunos expresan una ruptura más profunda del proceso de resolución que otros.

Para ilustrar el procedimiento de codificación, considérese un reactivo de trigonometría orientado al cálculo de una razón trigonométrica y su simplificación final. Se asignó $\text{ErrorCode}=0$ cuando el estudiante desarrolló correctamente el procedimiento y obtuvo el resultado esperado. Se asignó $\text{ErrorCode}=1$ cuando el método elegido fue adecuado, pero el desarrollo incluyó un error procedimental específico, por ejemplo, una transformación algebraica incorrecta o una simplificación inválida. Se asignó $\text{ErrorCode}=2$ cuando el procedimiento general fue correcto, pero el resultado final fue erróneo por un fallo de sustitución, signo o cálculo numérico. Se asignó $\text{ErrorCode}=3$ cuando la respuesta presentó un fallo total, es decir, un desarrollo incompatible con el planteamiento del reactivo o una secuencia de pasos que invalidó por completo la solución. Finalmente, se asignó $\text{ErrorCode}=4$ cuando la resolución quedó parcial o inconclusa, mostrando intención de respuesta sin completar los pasos necesarios para llegar a un resultado interpretable.

Encuesta de cierre

Se aplicó una encuesta post-curso ($N=154$) con escala Likert (1-5) y dominios: percepción global, retroalimentación, evaluaciones del curso, satisfacción global, hábitos de estudio y dificultades percibidas. Se estimó consistencia interna por dominio (α de Cronbach) para respaldar el uso de promedios por dominio como indicadores agregados.

A modo de ejemplo, en el dominio de percepción de aprendizaje se incluyeron reactivos tipo Likert como los siguientes: "Comprendí los procedimientos básicos asociados a la unidad evaluada", "Me sentí

capaz de resolver ejercicios similares a los revisados en clase" y "Identifiqué con claridad los errores cometidos después de la retroalimentación". En el dominio de retroalimentación se incorporaron reactivos como: "La retroalimentación recibida me ayudó a reconocer en qué parte del procedimiento fallé" y "Las explicaciones posteriores al examen me permitieron distinguir entre error de procedimiento y error de resultado". Estos reactivos se respondieron en una escala de 1 a 5, donde 1 correspondió al nivel más bajo de acuerdo y 5 al más alto.

Análisis

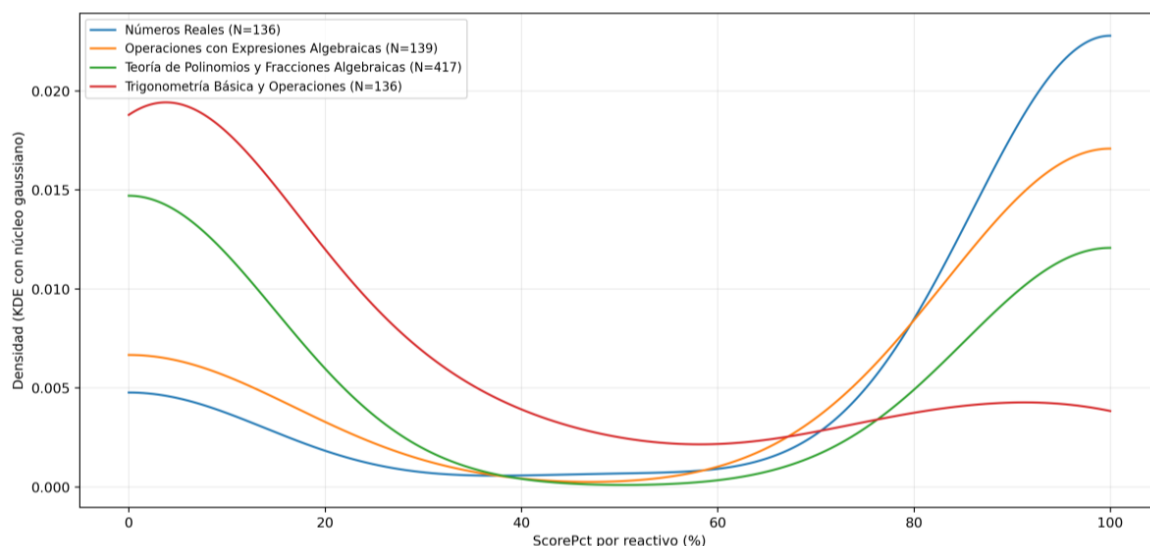
Se calcularon estadísticos descriptivos por unidad y por parcial, (i) ScorePct medio y dispersión, (ii) porcentaje de error severo, (iii) composición porcentual de ErrorCode por parcial. Para triangulación, se comparó la percepción media por unidad (Likert) con el desempeño objetivo por unidad (ScorePct). Las figuras reportadas incluyen distribuciones suavizadas (KDE gaussiana) y visualizaciones comparativas (barras y mapas de calor) para facilitar la lectura.

Resultados

Desempeño y severidad por unidad temática

La Tabla 1 resume desempeño (ScorePct) y severidad (error severo) por unidad. Se observa heterogeneidad marcada. Trigonometría Básica y Operaciones presenta el menor ScorePct medio (24.43%) y la mayor tasa de error severo (91.18%), configurando la unidad crítica del curso. En contraste, Números Reales exhibe el mejor desempeño (81.99%) y una severidad baja (19.12%). Unidades con alta carga procedimental y transición representacional, como Teoría de Polinomios y Fracciones Algebraicas ($\text{ScorePct}=45.08\%$; $\text{severidad}=54.92\%$) y Números Complejos ($\text{ScorePct}=58.21\%$; $\text{severidad}=48.53\%$), también muestran perfiles de riesgo elevados.

Figura 1.
Distribución del
 ScorePct por unidad
(KDE con núcleo
gaussiano)



Se comparan curvas de densidad del puntaje porcentual por reactivo (ScorePct) en un conjunto focal de unidades (Números Reales, Operaciones con Expresiones Algebraicas, Polinomios/Fracciones Algebraicas y Trigonometría). La figura evidencia diferencias marcadas en forma y ubicación de las distribuciones, distinguiendo unidades con concentración de valores altos frente a unidades con acumulación de puntajes bajos y colas pronunciadas. Este patrón respalda la hipótesis principal de heterogeneidad del desempeño por unidad.

De manera complementaria al análisis descriptivo, la asociación entre porcentaje de error severo y desempeño medio por unidad mostró una relación negativa alta, lo que reforzó la consistencia interna

del diagnóstico: a mayor severidad del error, menor ScorePct medio en la unidad correspondiente.

Evolución del perfil de error por parcial

La Tabla 2 reporta la composición porcentual de ErrorCode por parcial, usando el reactivo como unidad de observación. El segundo parcial presenta una alta proporción de fallos totales (48.20%) sin presencia de errores de procedimiento o resultado aislados (según la codificación registrada), mientras que el tercer parcial muestra un perfil más diversificado con errores de procedimiento (13.70%) y de resultado (27.64%), además de fallo total (32.69%). Este hallazgo es coherente con el hecho de que cada parcial evalúa contenidos distintos y, por tanto, expone distintos tipos de dificultad.

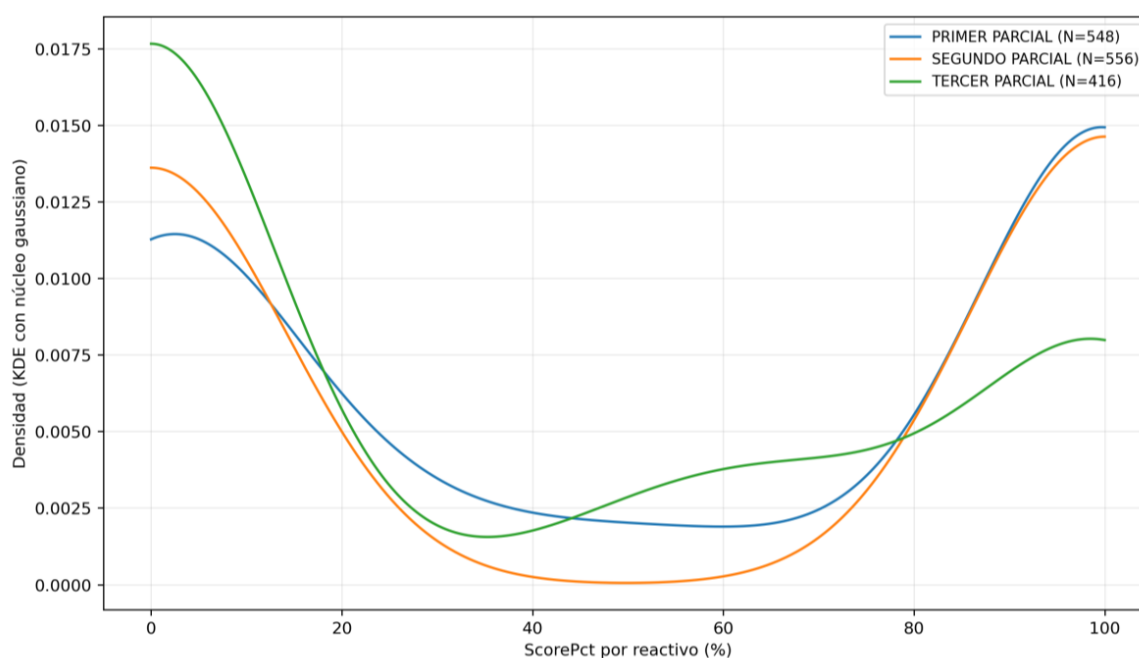


Figura 2. Distribución descriptiva del ScorePct por corte evaluativo (KDE con núcleo gaussiano)

Se presentan las curvas de densidad del ScorePct para cada corte evaluativo usando el reactivo como unidad de observación. Su lectura es estrictamente descriptiva, dado que cada parcial integró unidades temáticas distintas; por ello, la figura no debe interpretarse como una medición repetida del mismo contenido, sino como un contraste contextual de la distribución del rendimiento observado en cada examen parcial.

Encuesta: percepción, confiabilidad y brecha percepción-desempeño

La Tabla 3 resume dominios de la encuesta y consistencia interna (α). Los dominios muestran α

altos en percepción global (0.985), retroalimentación (0.939) y evaluaciones (0.904), lo que respalda su estabilidad interna. No obstante, al comparar percepción media por unidad con desempeño objetivo (Tabla 1), se observa una descalibración: las percepciones por unidad se mantienen relativamente altas (≈ 3.75 – 3.84), aun cuando el desempeño cae drásticamente en trigonometría. Este patrón sugiere que el estudiante puede “sentir” dominio sin lograr ejecución correcta, reforzando la pertinencia de intervención formativa y de estrategias orientadas a mejorar la correspondencia entre percepción de dominio y ejecución efectiva. (Pintrich, 2004).

Tabla 1. Desempeño, severidad y percepción por unidad temática

Unidad temática	N reactivos	N estudiantes	ScorePct media (%)	ScorePct SD	Error severo (%)	Percepción media (1–5)
Números Reales	136	136	81.99	37.85	19.12	3.84
Números Complejos	272	136	58.21	45.59	48.53	3.75
Operaciones con Expresiones Algebraicas	139	139	71.94	45.09	28.06	3.82
Teoría de Polinomios y Fracciones Algebraicas	417	139	45.08	49.82	54.92	3.78
Ecuaciones de Grado Superior	420	105	35.31	43.84	34.29	3.83
Trigonometría Básica y Operaciones	136	136	24.43	34.25	91.18	3.79

Parcial	N_total	Sin error (%)	Proc. (%)	Res. (%)	Fallo total (%)	Parcial/Incompleto (%)
PRIMER PARCIAL	548	47.81	0.00	0.00	30.47	21.72
SEGUNDO PARCIAL	556	51.80	0.00	0.00	48.20	0.00
TERCER PARCIAL	416	25.00	13.70	27.64	32.69	0.96

Tabla 2. Composición porcentual del tipo de error por parcial (unidad de observación: reactivo)

Conviene precisar que, en el primer y segundo parcial, la codificación registrada concentró los fallos en las categorías de “fallo total” (3) y “parcial/incompleto” (4), por lo que no se documentaron casos clasificados de manera aislada como “error de procedimiento” (1) o “error de re-

sultado” (2). Este comportamiento respondió al criterio de codificación aplicado durante el vaciado de los reactivos en esos cortes, donde predominó la clasificación del fallo dominante cuando el desarrollo del estudiante comprometía de forma integral la validez de la respuesta.

Tabla 3. Dominios de la encuesta de cierre: media, dispersión y confiabilidad interna

Dominio (encuesta)	k ítems	α Cronbach	N	Media (1–5)	SD
Percepción global	8	0.98	154	3.81	1.35
Retroalimentación	6	0.94	154	3.86	1.12
Evaluaciones del curso	6	0.90	154	3.85	1.03
Satisfacción global	4	0.96	154	3.97	1.19
Hábitos de estudio	6	0.73	154	3.40	0.84
Dificultades percibidas	8	0.93	154	3.02	1.16

Discusión

Los resultados confirman que el desempeño en Matemáticas Básicas de Ingeniería es altamente dependiente del contenido y del tipo de habilidades exigidas (Hattie & Timperley, 2007). El caso de trigonometría —bajo ScorePct y severidad extrema— sugiere un déficit estructural de ejecución (procedimiento, control de signos, uso de identidades, verificación por cuadrante/ángulo), más que un problema superficial de memoria. En contraste, Números Reales muestra que una parte de la cohorte posee bases funcionales, aunque persiste un subgrupo con fragilidades (Boud & Molloy, 2013).

Desde la teoría del error, la tipología por reactivo permite pasar de la calificación global a un diagnóstico accionable: qué contenidos concentran error severo, y por tanto dónde conviene invertir tiempo instruccional adicional. La encuesta aporta un hallazgo complementario: la percepción no se alinea plenamente con la ejecución, especialmente en unidades críticas, lo que es consistente con fenómenos de posible desajuste entre percepción y ejecución (Nicol & Macfarlane-Dick, 2006). En términos didácticos, ello implica que no basta con “explicar” o “repasar”; se requiere diseñar práctica distribuida y

retroalimentación inmediata que haga visible el error, reduzca el costo cognitivo de corregirlo y obligue a verificar pasos clave.

En IME, la relevancia disciplinar de trigonometría, complejos y polinomios es transversal (fasores, señales, modelado y análisis). Por tanto, atender estas unidades no solo mejora el desempeño del curso, sino que puede reducir rezago en asignaturas posteriores. Este artículo propone que la matriz por reactivo se institucionalice como herramienta mínima de mejora continua: con captura semestral consistente, se generan series históricas que permiten medir impacto de intervenciones (Felder & Brent, 2005) y construir evidencia publicable.

Conclusiones y recomendaciones

El análisis a nivel de reactivo evidenció una heterogeneidad sustantiva entre unidades temáticas en términos de desempeño (ScorePct) y severidad del error, bajo un criterio homogéneo de clasificación. En particular, la unidad de Trigonometría Básica y Operaciones se identificó como el foco crítico del curso al concentrar simultáneamente el menor desempeño promedio y la mayor proporción de errores severos. Asimismo, la composición del tipo de error mostró variaciones entre parciales; dichas variaciones resultaron consistentes con la estructura real del curso, dado que cada parcial evaluó conjuntos de contenidos distintos y, por tanto, expuso perfiles de dificultad diferentes.

La triangulación entre evidencia objetiva (exámenes) y evidencia perceptual (encuesta de cierre) indicó que la percepción reportada por los estudiantes no se correspondió plenamente con la ejecución observada en unidades críticas, lo que sugirió la presencia de brechas entre la percepción reportada y la ejecución observada, particularmente en las unidades críticas. En conjunto, estos hallazgos respaldaron la utilidad de transformar la calificación global en un diagnóstico accionable por unidad y por patrón de error, con implicaciones directas para la mejora didáctica en cursos habilitantes de ingeniería.

A partir de los datos obtenidos, se recomendó priorizar en primer término la unidad de Trigonometría Básica y Operaciones, dado que concentró la mayor proporción de error severo (91.18%) y el menor desempeño promedio del curso. En segundo término, se sugirió atender Teoría de Polinomios y Fracciones Algebraicas y Números Complejos mediante estrategias formativas de alta frecuencia, tales como microevaluaciones semanales, retroalimentación inmediata, listas de verificación de procedimiento y verificación por pasos críticos, con el objetivo de reducir la severidad del error y mejorar la estabilidad operativa del estudiante.

Para fortalecer la validez inferencial de los resultados, se sugirió desarrollar estudios posteriores con diseños pre-post o cuasi experimentales por grupo, así como replicaciones por cohorte que permitan estimar

estabilidad de patrones, tamaños de efecto y sensibilidad a la selección de reactivos.

Referencias

- Biggs, J., & Tang, C. (2011). Teaching for quality learning at university (4th ed.). Open University Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698-712.
- Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE.
- Felder, R. M., & Brent, R. (2005). Understanding student differences. *Journal of Engineering Education*, 94(1), 57-72.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, M. d. P. (2022). Metodología de la investigación (7a ed.). McGraw-Hill.
- Movshovitz-Hadar, N., Zaslavsky, O., & Inbar, S. (1987). An empirical classification model for errors in high school mathematics. *Journal for Research in Mathematics Education*, 18(1), 3-14.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16, 385-407.
- Prince, M. J., & Felder, R. M. (2006). Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Journal of Engineering Education*, 95(2), 123-138.
- Radatz, H. (1979). Error analysis in mathematics education. *Journal for Research in Mathematics Education*, 10(3), 163-172.
- Rosado Capetillo, A., Zapata Rodríguez, U. G., & Ortiz Martínez, F. (2025). Mejora continua en la Facultad de Ingeniería Mecánica y Ciencias Navales de la Universidad Veracruzana. *Acreditas*, 17. <https://doi.org/10.61752/acd.i17.263>
- Schoenfeld, A. H. (1985). Mathematical problem solving. Academic Press.
- Tall, D. (2013). How humans learn to think mathematically: Exploring the three worlds of mathematics. Cambridge University Press.
- Tinto, V. (2017). Through the eyes of students. *Journal of College Student Retention: Research, Theory & Practice*, 19(3), 254-269.
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64-70.