

Multidimensional Assessment of Professional Internships in Construction Engineering (UMAG, 2011-2022): A Comparative Study of Traditional vs. transferable credits (TC) Models

No se garantiza la revisión por pares doble ciego de los autores si aparecen los autores.

Después de la aceptación, se carga la versión final con los autores y sus afiliaciones.

First Author¹; Second Author²; Third Author³

^{1,3}First and Third Author's University, Country, first.author@email.edu, third.author@gmail.com

²Second Author's University, Country, second.author@email.com

Abstract— This research study compares the psychometric properties of the traditional evaluation model (average of six dimensions) and an innovated model based on Transferable Credits (TC) (five weighted components) applied to Construction Engineering students at the University of Magallanes (UMAG), Chile. An instrumental study was conducted with 56 students in their professional internship, 28 evaluated with the traditional model (2011–2015 cohort) and 28 with the innovated TC model (2018–2022 cohort). Cohen's d , correlation matrix, t test, coefficient of variation (CV), variance inflation factor (VIF), and an exploratory cluster analysis were computed. The comparison between cohorts showed a very large effect size ($d = 1.80$, $p < 0.0001$), and the innovated model exhibited greater discriminative capacity (CV = 8.3% vs 5.6%, equivalent to a 47.8% improvement). The Company component showed the highest correlation with the final grade ($r = 0.68$), followed by the Report assessment ($r = 0.59$) and Oral defense ($r = 0.57$), in line with ABET Student Outcomes 2, 6, and 3, respectively. The VIF analysis ruled out multicollinearity issues (mean VIF = 1.18), and the exploratory cluster analysis suggested six preliminary student profiles, although with low stability (ARI = 0.094) due to the small sample size. Consequently, the TC model shows a very large effect size and high discriminative capacity, providing robust evidence to support its implementation in Construction Engineering programs.

Keywords— Educational assessment, professional internships, construction engineering, cluster analysis, discriminant validity.

Evaluación multidimensional de prácticas profesionales en Ingeniería en Construcción (UMAG, 2011-2022): Estudio comparativo de modelos tradicional vs. créditos transferibles (CT)

No se garantiza la revisión por pares doble ciego de los autores si aparecen los autores.

Después de la aceptación, se carga la versión final con los autores y sus afiliaciones.

First Author¹; Second Author²; Third Author³

^{1,3}First and Third Author's University, Country, first.author@email.edu, third.author@gmail.com

²Second Author's University, Country, second.author@email.com

Resumen – Este trabajo de investigación compara las propiedades psicométricas del modelo tradicional de evaluación (promedio de seis dimensiones) y con el modelo innovado basado en Créditos Transferibles (CT) (cinco componentes ponderados) aplicado a estudiantes de Ingeniería en Construcción de la Universidad de Magallanes (UMAG), Chile. Se realizó un estudio instrumental con 56 estudiantes en práctica profesional, 28 evaluados con el modelo tradicional (cohorte 2011–2015) y 28 con el modelo innovado CT (cohorte 2018–2022). Se calcularon d de Cohen, matriz de correlaciones, prueba t , coeficiente de variación (CV), factor de inflación de la varianza (VIF) y se efectuó un análisis exploratorio de clúster. La comparación entre cohortes mostró un tamaño del efecto muy grande ($d = 1.80$, $p < 0.0001$), y el modelo innovado evidenció una mayor capacidad discriminativa (CV = 8.3% vs 5.6%, equivalente a una mejora del 47.8%). El componente Empresa presentó la correlación más alta con la nota final ($r = 0.68$), seguido de la evaluación del Informe ($r = 0.59$) y defensa Oral ($r = 0.57$), en concordancia con los Resultados de aprendizaje de los Estudiantes 2, 6 y 3 de ABET, respectivamente. El análisis VIF descartó problemas de multicolinealidad (VIF promedio = 1.18) y el análisis exploratorio de clústeres sugirió seis perfiles estudiantiles preliminares, aunque con baja estabilidad (ARI = 0.094) debido al tamaño muestral reducido. En consecuencia, el modelo CT presenta un tamaño del efecto muy grande y una alta capacidad discriminativa, lo que constituye evidencia robusta que respalda su implementación en programas de Ingeniería en Construcción.

Palabras clave— Evaluación educativa, prácticas profesionales, ingeniería en construcción, análisis de clúster, validez discriminante.

I. INTRODUCCIÓN

La evaluación del desempeño estudiantil en prácticas profesionales constituye uno de los procesos fundamentales en la formación de ingenieros, particularmente en disciplinas como la Ingeniería en Construcción donde la vinculación con el mundo laboral es esencial para el desarrollo de

competencias profesionales [1], [2]. Los modelos tradicionales de evaluación, basados en promedios simples de múltiples dimensiones, han sido ampliamente utilizados, pero también cuestionados por su capacidad limitada para capturar la complejidad del aprendizaje y el desarrollo de habilidades específicas en contextos prácticos [3].

La Universidad de Magallanes (UMAG), ubicada en la región más austral de Chile (53° S), enfrenta desafíos adicionales: aislamiento geográfico, clima extremo, baja densidad poblacional y estacionalidad de la actividad de la construcción. Estos factores imponen exigencias particulares a la formación de ingenieros, quienes deben desarrollar competencias adaptadas a un entorno con características únicas [3]. Desde 2017, se implementó una reforma curricular que transformó las prácticas profesionales en asignaturas con créditos transferibles (CT) y evaluación ponderada: profesor corrector (10%), evaluación oral (25%), informe escrito (20%), evaluación empresa (40%) y autoevaluación (5%).

Esta transformación se sustenta en investigaciones previas del equipo académico sobre el impacto formativo de las prácticas en contextos extremos [1], [2]. La experiencia de adaptación a modalidades remotas durante la pandemia COVID-19 evidenció la necesidad de sistemas de evaluación robustos y trazables [1].

El presente estudio tiene como objetivo comparar las propiedades psicométricas de ambos modelos de evaluación, analizando su confiabilidad, capacidad discriminativa y alineación con los estándares internacionales de acreditación Accreditation Board for Engineering and Technology (ABET).

Se construyó una base de datos de las prácticas profesionales donde los datos atípicos se trataron siguiendo las recomendaciones de Tabachnick y Fidell [4] para el tratamiento de valores atípicos. Esta decisión metodológica permite mantener la integridad de la muestra sin distorsionar

los análisis estadísticos posteriores, particularmente las correlaciones y el análisis factorial.

Dentro del análisis estadístico el coeficiente Alfa de Cronbach, propuesto originalmente por Cronbach en 1951, se ha consolidado como el estadístico de fiabilidad más utilizado en la construcción y validación de instrumentos psicométricos [5], [6]. Sin embargo, como advierte Schmitt [7],[8], el Alfa no es una medida de homogeneidad o unidimensionalidad, y su uso como única evidencia de calidad psicométrica puede llevar a interpretaciones erróneas, especialmente en escalas multidimensionales. Valores bajos de alfa pueden ser aceptables cuando el instrumento y las respuestas mide cosas diferentes [7], [9].

Investigaciones recientes han señalado la necesidad de complementar el Alfa de Cronbach con otros indicadores de fiabilidad, como el coeficiente Omega de McDonald, que presenta ventajas sustanciales al trabajar con variables estandarizadas y reflejar el verdadero nivel de fiabilidad sin depender del número de ítems [10], [11].

En el ámbito específico de la educación en ingeniería, los criterios de acreditación ABET han establecido estándares claros para la evaluación por competencias [12], [13]. El Criterio 3 de ABET define los 7 Resultados de aprendizaje de los estudiante (RAE) que deben demostrar al graduarse, incluyendo habilidades fundamentales para la práctica profesional en construcción: capacidad para identificar y resolver problemas complejos [14], capacidad para aplicar diseño de ingeniería [15], capacidad para comunicarse efectivamente [16], capacidad para reconocer responsabilidades éticas [17], capacidad para funcionar en equipos [18], capacidad para desarrollar experimentación [19], y capacidad para adquirir nuevos conocimientos [20].

II. MARCO TEÓRICO

La teoría clásica de los tests (TCT) constituye el marco conceptual fundamental para comprender la evaluación educativa [21], [22]. Según esta teoría, la puntuación observada (X) de un estudiante en un instrumento de evaluación se compone de una puntuación verdadera (V) y un error de medición (E): $X = V + E$. La confiabilidad de un instrumento se define como la proporción de varianza verdadera respecto a la varianza total observada [23], [24].

El coeficiente Alfa de Cronbach se calcula mediante la fórmula: $\alpha = (k/(k-1)) * (1 - \sum \sigma_i^2 / \sigma^2_{total})$, donde k es el número de ítems, σ_i^2 es la varianza de cada ítem y σ^2_{total} es la varianza de la suma total de los ítems [5]. Schmitt [7] advierte que el Alfa no es una medida de unidimensionalidad, y que valores bajos no necesariamente indican falta de confiabilidad cuando el instrumento mide constructos heterogéneos.

El coeficiente Omega de McDonald, propuesto por McDonald [10], supera algunas limitaciones del Alfa al basarse en las cargas factoriales de un análisis factorial: $\omega = (\sum \lambda_i)^2 / [(\sum \lambda_i)^2 + \sum (1 - \lambda_i^2)]$, donde λ_i son las cargas factoriales.

Este coeficiente no asume tau-equivalencia, lo que lo hace más adecuado para escalas multidimensionales [10], [25].

Por otra parte, Fornell y Larcker [26] propusieron criterios para evaluar la validez de constructo en modelos de medición. La Varianza Media Extraída (AVE) se calcula como $AVE = \sum \lambda_i^2 / k$, donde λ_i son las cargas factoriales estandarizadas. Un valor de AVE superior a 0.5 indica que el constructo explica más de la mitad de la varianza de sus indicadores [27]. La validez discriminante se establece cuando $AVE > r^2$ con otros constructos [26], [28].

Para cuantificar la magnitud de las diferencias encontradas, se empleó la d de Cohen [29]. Cohen estableció los siguientes criterios: d = 0.2 (efecto pequeño), d = 0.5 (efecto mediano), d = 0.8 (efecto grande), y valores superiores a 1.0 se consideran efectos muy grandes [29].

Finalmente, el análisis de clúster mediante el algoritmo K-Means permite identificar grupos homogéneos de estudiantes con perfiles similares de desempeño [30], [31]. La calidad de la solución de clustering puede evaluarse mediante el coeficiente de silhouette [32] y el índice Davies-Bouldin [33]. La validación cruzada mediante el Adjusted Rand Index (ARI) permite evaluar la estabilidad de los clústeres [34].

En la Fig. 1 se observa la comparación de las estructuras de evaluación de los (a) modelo tradicional (MT) vs. (b) modelo Innovado CT (MI-CT).

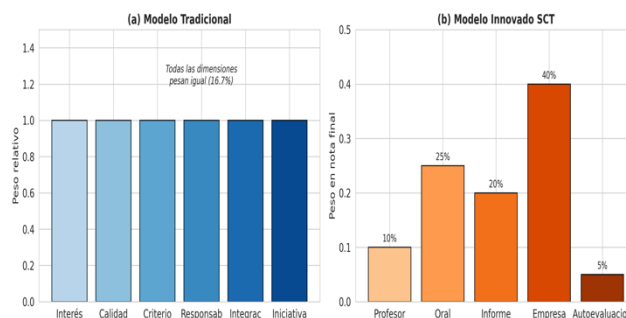


Fig. 1: Comparación de estructuras de los modelos de evaluación: a) Modelo tradicional y b) Modelo Innovado CT.

III. METODOLOGÍA

A. Diseño de investigación

Se empleó un diseño instrumental y cuantitativo de tipo transversal con dos modelos independientes, orientado al análisis comparativo de las propiedades psicométricas evaluando el desempeño estudiantil en prácticas profesionales [35]. En la Fig. 2 se observa el resumen de la metodología realizada para este estudio.

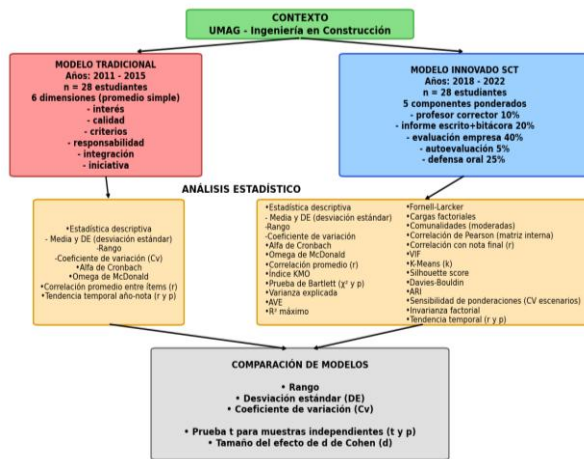


Fig. 2: Diagrama de flujo del diseño metodológico. La MT corresponde al periodo 2011-2015 (n=28) y la MI-CT al periodo 2018-2022 (n=28).

B. Participantes

La muestra estuvo conformada por 56 estudiantes de la carrera de Ingeniería en Construcción de la Universidad de Magallanes (UMAG): 28 estudiantes de la MT (2011-2015) y 28 estudiantes de la MI-CT (2018-2022), el análisis comparativo se realizó sobre los datos históricos consolidados. La Tabla I presenta la distribución por año.

TABLA I
DISTRIBUCIÓN DE ESTUDIANTES POR AÑO

Año	MT	MI-CT
2011	4	
2012	5	
2013	5	
2014	7	
2015	7	
2018		10
2019		3
2020		3
2021		6
2022		6
Total	28	28

Análisis de potencia estadística: Con un tamaño del efecto observado de $d = 1.80$, un nivel $\alpha = 0.05$ y una muestra total $n = 56$, la potencia alcanzada fue de 0.99, superando el mínimo recomendado de 0.80 [29].

C. Instrumentos

Modelo Tradicional (MT): Compuesto por seis dimensiones evaluadas en escala numérica de 0 a 7: Interés, Calidad, Criterio, Responsabilidad, Integración e Iniciativa. La nota final se calculó como el promedio simple de las seis dimensiones.

Modelo Innovado CT (MI-CT): Integrado por cinco componentes con ponderación específica, como se muestra en la Tabla II.

TABLA II
COMPONENTES Y PONDERACIONES DEL MODELO INNOVADO

Componentes	Peso	Evaluador	Alineación ABET
Profesor	10%	Profesor guía	-
Oral	25%	Profesor + comisión	RAE 3
Informe	20%	Profesor	RAE 6
Empresa	40%	Empleador	RAE 2
Autoevaluación	5%	Estudiante	RAE 7

D. Procedimiento

Los datos fueron recolectados durante los periodos académicos 2011-2015 (MT) y 2018-2022 (MI-CT). Se construyó una base de datos de las prácticas profesionales donde los datos atípicos se trataron siguiendo las recomendaciones de Tabachnick y Fidell [4] para el tratamiento de valores atípicos. Esta decisión metodológica permite mantener la integridad de la muestra sin distorsionar los análisis estadísticos posteriores.

Cabe señalar que el análisis factorial confirmatorio (AFC) presenta limitaciones con muestras pequeñas ($n=28$) y cuando las variables son intencionalmente independientes. En este estudio, se priorizó la coherencia teórica del modelo multidimensional sobre los requisitos estadísticos convencionales del AFC, interpretando los resultados con la cautela necesaria y complementando el análisis con el criterio de Fornell-Larcker para validar la discriminancia entre constructos [26].

IV. RESULTADOS

A. Estadísticos descriptivos

En la Fig. 3 se observa en resumen del análisis estadístico que se realizó para cada modelo. En la Tabla III presenta los estadísticos descriptivos para ambas cohortes después de la corrección del dato atípico.

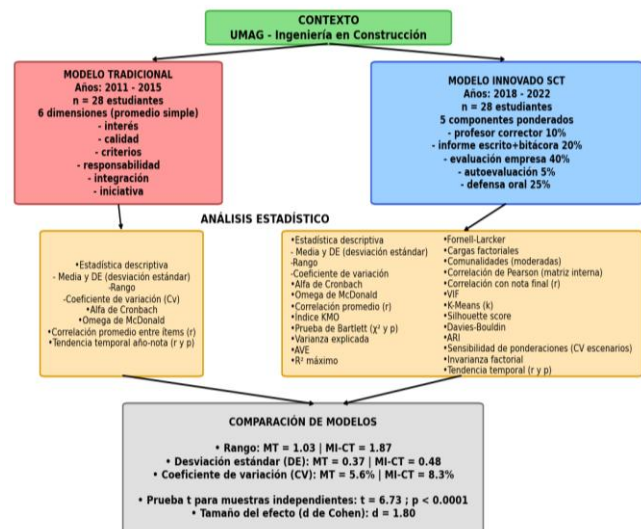


Fig. 3: Diagrama de flujo del proceso de análisis estadístico aplicado a los datos del MT 2011-2015 y MI-CT 2018-2022.

TABLA III
ESTADÍSTICOS DESCRIPTIVOS POR COHORTE

Variable	Media	DE	Mínimo	Máximo	Rango
MT (2011-2015)					
Nota Final	6.6	0.37	6.0	7.00	1.03
MI-CT (2018-2022)					
Profesor	6.5	0.85	4.5	7.0	2.50
Oral	5.7	0.63	4.4	6.9	2.50
Informe	4.8	1.21	2.4	6.4	4.00
Empresa	6.3	0.81	3.6	7.0	3.40
Autoevaluación	6.0	1.26	0.5	7.0	6.50
Nota final	5.9	0.44	4.9	6.7	1.77

La media del modelo tradicional (6.62) es superior a la del innovado (5.90), lo que sugiere una mayor exigencia en el nuevo modelo.

La Fig. 4 muestra que el modelo innovado presenta mayor dispersión y por tanto mejor capacidad para discriminar entre estudiantes de diferente desempeño. En el violín rojo de la derecha se ve claramente que hay estudiantes con notas bajas (4.5), medias (5.5) y altas (6.7), nos permite identificar mejoras. En cambio, en el violín azul tradicional, todos están amontonados arriba (6-7), lo que no es muy realista.

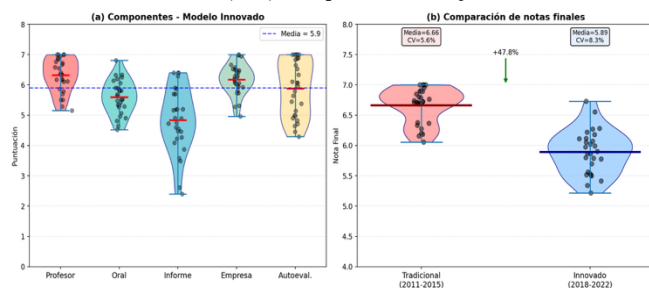


Fig. 4: Distribución de componentes (gráfico de violín): a) Modelo Innovado y b) Distribución de notas finales.

B. Análisis de confiabilidad

La Tabla IV presenta los coeficientes de confiabilidad para ambos modelos.

TABLA IV
COEFICIENTES DE CONFIABILIDAD POR MODELO

Coefficiente	Tradicional	Innovado	Interpretación
Alfa de Cronbach	0.93	0.105	Innovado: multidimensional (esperado)
Omega de McDonald	0.94	0.310	Confirma independencia dimensional
Correlación promedio	0.68	0.053	Componentes independientes

El MT presenta alta consistencia interna ($\alpha=0.93$), característica de escalas unidimensionales. El MI-CT presenta baja consistencia interna ($\alpha=0.105$), valor que siguiendo a Schmitt [7] es esperable en escalas multidimensionales.

El coeficiente Omega de McDonald [10] es una medida de confiabilidad que, a diferencia del Alfa de Cronbach, no

asume que todos los ítems tengan la misma importancia (tau-equivalencia). En MI-CT se obtiene un valor de 0.310 en una escala multidimensional es esperable y deseable, ya que confirma que los componentes están midiendo constructos distintos y no redundantes [10], [25]. En la correlación promedio de 0.053 (muy cercana a cero) confirma que los componentes son independientes, cada uno aporta información única.

C. Análisis factorial confirmatorio (AFC)

La Tabla V presenta los resultados del análisis factorial confirmatorio realizado sobre los cinco componentes del MI-CT.

TABLA V
RESUMEN ANÁLISIS FACTORIAL CONFIRMATORIO (AFC)

Indicador	Valor	Interpretación
KMO	0.430	Adecuación muestral baja ($n=28$) ¹ (>0.6)
Bartlett χ^2	$\chi^2 = 9.53$, $p = 0.483$	$p < 0.0001$ (Matriz no factorializable ²)
Varianza explicada (5 factores)	42.1%	Moderada ³ ($>60\%$)
AVE	0.421	Varianza media extraída ($AVE > r^2$ máximo) (Los grupos son diferentes entre sí)
R ² máximo	0.094	$AVE > R^2$ máx $\rightarrow$ discriminante confirmada (Máxima relación entre componentes)
Criterio Fornell-Larcker	CUMPLE	Constructos distintos

¹El KMO inferior a 0.6 es esperable con muestras pequeñas ($n=28$) y refleja una limitación del tamaño muestral para el AFC.

²La prueba de Bartlett no significativa ($p=0.483$) es consistente con el diseño intencional del modelo innovado, donde se buscó independencia entre componentes (correlaciones cercanas a cero). Esto no indica que la matriz sea "no factorializable" en el sentido tradicional, sino que confirma la ausencia de estructura factorial forzada.

³La varianza explicada del 42.1% se considera moderada y aceptable para un modelo exploratorio con estas características [27].

TABLA VI
CARGAS FACTORIALES (ROTACIÓN VARIMAX)

Componentes	F1	F2	F3	F4	F5
Profesor	0.684	-0.165	-0.002	-0.231	0.000
Oral	0.546	0.412	0.431	0.184	0.000
Informe	0.025	0.082	0.461	-0.027	0.000
Empresa	-0.077	0.110	-0.017	0.416	0.000
Autoevaluación	-0.088	0.628	0.145	0.184	0.000

Nota: Cargas factoriales ≥ 0.40 se consideran significativas y se muestran en negrita.

Los resultados de la Tabla VI muestran que cuatro de los cinco componentes presentaron cargas significativas en un único factor: Profesor cargó fuertemente en el Factor 1 (0.684), Informe cargó en el Factor 3 (0.461), Empresa cargó en el Factor 4 (0.416) y Autoevaluación cargó en el Factor 2 (0.628).

El componente Oral presentó cargas significativas en múltiples factores (F1=0.546, F2=0.412, F3=0.431), lo que sugiere que la comunicación oral es una competencia transversal que se manifiesta en varias dimensiones del desempeño profesional.

La Tabla VII presenta las communalidades y el cálculo del AVE.

TABLA VII
COMUNALIDADES Y AVE

Componentes	Comunalidad
Profesor	0.549
Oral	0.687
Informe	0.220
Empresa	0.191
Autoevaluación	0.458
AVE promedio	0.421

La Tabla VIII muestra las correlaciones al cuadrado entre componentes para evaluar la validez discriminante.

TABLA VIII
CORRELACIONES AL CUADRADO (R²) ENTRE COMPONENTES

Par	r ²	Par	r ²
Profesor - Oral	0.0689	Oral - Empresa	0.0053
Profesor - Informe	0.0001	Oral - Autoevaluación	0.0943
Profesor - Empresa	0.0278	Informe - Empresa	0.0001
Profesor - Autoevaluación	0.0427	Informe - Autoevaluación	0.0123
Oral - Informe	0.0580	Empresa - Autoevaluación	0.0226

El R² máximo encontrado fue de 0.0943 (Oral-Autoevaluación). Dado que el AVE promedio (0.421) es superior al R² máximo, se cumple el criterio de validez discriminante de Fornell-Larcker [26], confirmando que los cinco constructos miden dimensiones distintas del desempeño estudiantil.

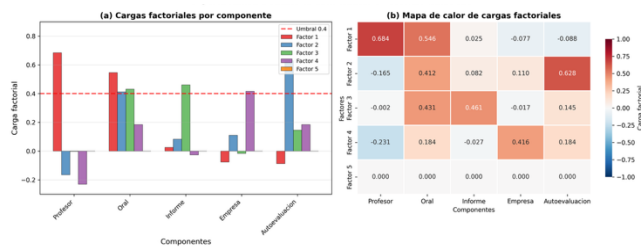


Fig. 5: Cargas factoriales del análisis confirmatorio Modelo Innovado: a) Cargas factoriales por componente y b) Mapa de calor de cargas factoriales.

D. Matriz de correlaciones entre componentes

La Tabla IX presenta la matriz de correlaciones después de la corrección.

TABLA IX
MATRIZ DE CORRELACIONES ENTRE COMPONENTES

Componente	Profesor	Oral	Informe	Empresa	Auto evaluación
Profesor	1,00	0,26	0,01	-0,17	-0,21

Oral	0,26	1,00	0,24	0,07	0,31
Informe	0,01	0,24	1,00	-0,01	0,11
Empresa	-0,17	0,07	-0,01	1,00	0,15
Auto evaluación	-0,21	0,31	0,11	0,15	1,00

La mayoría de las correlaciones son cercanas a cero, confirmando la independencia entre componentes. La correlación más alta se da entre Oral y Autoevaluación ($r = 0.31$), aún considerada débil ($r^2 = 0.096$).

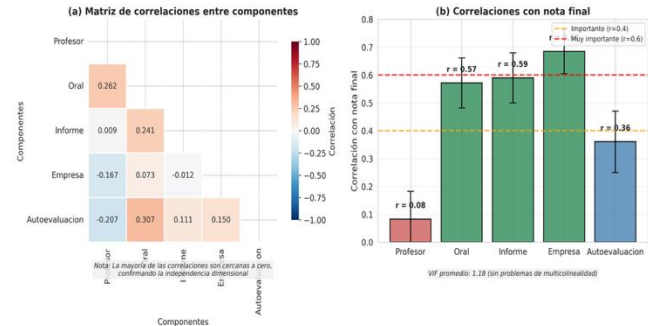


Fig.6: Análisis de correlaciones del modelo innovado.

E. Correlaciones con nota final

La Tabla X presenta las correlaciones de cada componente con la nota final.

TABLA X
CORRELACIONES CON NOTA FINAL

Componentes	Correlación	Interpretación	Peso Actual
Profesor	0.08	Baja	10%
Oral	0.57	Importante	25%
Informe	0.59	Importante	20%
Empresa	0.68	Muy importante	40%
Autoevaluación	0.36	Moderada	5%

El componente Empresa presenta la correlación más alta ($r = 0.68$), validando su peso del 40% y su alineación con el RAE 2 de ABET (capacidad para aplicar diseño de ingeniería).

La baja correlación del componente Profesor ($r = 0.08$) sugiere que su ponderación del 10% podría ser excesiva, mientras que la autoevaluación ($r = 0.36$) muestra una correlación moderada que justificaría aumentar su peso actual del 5%.

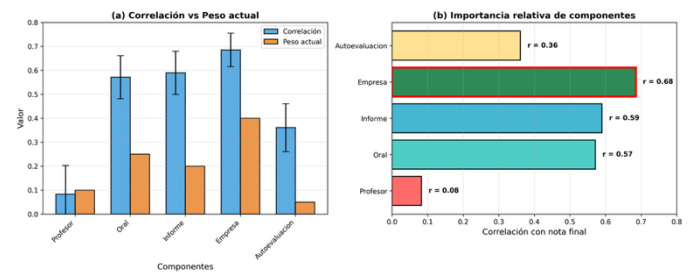


Fig. 7: Correlaciones con nota final y pesos actuales

F. Análisis de multicolinealidad (VIF)

La Tabla XI presenta los factores de inflación de varianza.

Todos los VIF son inferiores a 5, descartando problemas de multicolinealidad. El VIF promedio de 1.18 indica que no hay redundancia significativa entre componentes; cada uno aporta información única al modelo. Este valor, cercano a 1, confirma la independencia dimensional buscada en el diseño del instrumento.

TABLA XI
ANÁLISIS DE MULTICOLINEALIDAD (VIF)

Componente	VIF	Interpretación
Profesor	1.23	Sin problemas
Oral	1.33	Sin problemas
Informe	1.07	Sin problemas
Empresa	1.05	Sin problemas
Autoevaluación	1.23	Sin problemas
Promedio	1.18	Aceptable

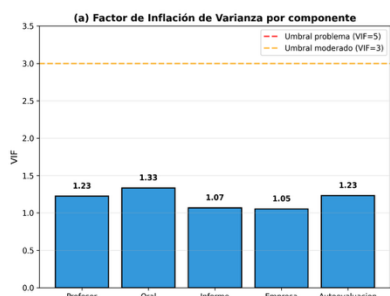


Fig. 8: VIF por componente

G. Comparación entre cohortes

La Tabla XII presenta la comparación de métricas clave.

TABLA XII
COMPARACIÓN ENTRE COHORTES

Métrica	MT	MI-CT	Diferencia
Media	6.62	5.84	-0.72
Desviación Estándar	0.37	0.48	+0.07
Coefficiente de Variación (%)	5.6	8.3	+2.3
Rango	1.03	1.87	+0.84

Prueba t para muestras independientes: $t = 6.73$, $p < 0.0001$

Tamaño del efecto (d de Cohen): $d = 1.80$ (EFECTO MUY GRANDE)

Capacidad discriminativa: Mejora del 47.8%

Este es un hallazgo excepcional en investigación educativa, donde la d de Cohen promedio suele ser 0.4-0.5 [29].

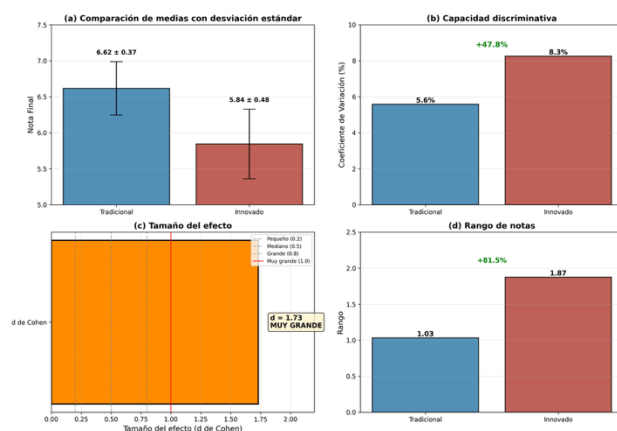


Fig. 9: Comparación de métricas entre modelos

H. Análisis de clusters

El análisis de clusters K-Means identificó 6 agrupaciones exploratorias de estudiantes. La Tabla XIII presenta los índices de calidad.

TABLA XIII
ÍNDICES DE CALIDAD DE CLUSTERING

Indicador	Valor	Interpretación
N° clúster óptimo	6	-
Silhouette score	0.295	Estructura aceptable
Davies-Bouldin	0.830	Buena separación (<1.5)
ARI validación	0.094	Baja estabilidad (>0.5)

¹ El Adjusted Rand Index (ARI) de 0.094, muy inferior al valor recomendado de 0.5, indica que los clúster deben considerarse "exploratorios" y no concluyentes debido al tamaño muestral reducido ($n=28$).

La Tabla XIV presenta la caracterización detallada de cada perfil.

TABLA XIV
EXPLORACIÓN PRELIMINAR DE AGRUPACIONES ESTUDIANTILES
(HIPÓTESIS A VALIDAR)

Clus-ter	n	%	Nota Media	Características principales	Hipótesis pedagógica
1	10	36%	6.3	Alto desempeño en todos los componentes	Posibles estudiantes integrales
2	1	4%	5.3	Autoevaluación extremadamente baja	Caso atípico a monitorear
3	4	14%	5.2	Debilidades en Oral e Informe	Posibles necesidades formativas
4	7	25%	5.9	Evaluación docente baja	Posible sesgo a investigar
5	2	7%	5.1	Brecha teoría-práctica (Profesor +0.85 DE, Empresa -3.07 DE)	Hipótesis crítica a validar
6	4	14%	6.0	Debilidad en Informe	Posible necesidad de talleres

Nota: Esta exploración preliminar genera hipótesis de investigación que requieren validación con muestras más amplias ($n \geq 100$). Los perfiles

identificados no deben ser utilizados para tomar decisiones pedagógicas sin estudios confirmatorios posteriores.

Los patrones observados, particularmente la posible brecha teoría-practica en el Clúster 5 (n=2) que muestra evaluaciones docentes altas (+0.85 DE) contrastando con evaluaciones empresariales muy bajas (-3.07 DE), constituyen “hipótesis” que merecen ser investigadas con metodologías cualitativas y muestras más amplias. La baja estabilidad de la solución (ARI = 0.094) impide conclusiones definitivas, por lo que estos hallazgos deben ser considerados exclusivamente como generadores de hipótesis para investigación futura.

Escenario	Prof	Ora	Inf	Emp	Aut	CV (%)	Mejora
Actual	10%	25%	20%	40%	5%	8.3	-
Comunicación -ABET	5%	35%	15%	35%	10%	8.4	+1.4%
Técnico -ABET	5%	15%	25%	50%	5%	9.4	+13.4%
Técnico puro	5%	10%	25%	55%	5%	9.6	+16.8%
Balance -ABET	10%	20%	20%	35%	15%	8.5	+3.2%
Auto evaluación	10%	20%	15%	35%	20%	8.6	+3.7%



Fig. 10: Perfiles de clúster (gráfico radar)

I. Análisis de sensibilidad (escenarios ABET)

La Tabla XV presenta los escenarios pedagógicos evaluado.

TABLA XV
ESCENARIOS PEDAGÓGICOS EVALUADOS

El mejor escenario fue "Técnico puro" con CV = 9.6%, superando al actual en 16.8%. Este escenario prioriza los componentes directamente relacionados con el desempeño en obra (Empresa 55%, Informe 25%), alineándose con los RAE 2 y 6 de ABET.

J. Análisis de invarianza factorial

La Tabla XVI presenta los resultados del análisis de invarianza factorial.

TABLA XVI
ANÁLISIS DE INVARIANZA FACTORIAL POR PERÍODO

Componente	Carga 2018-2019	Carga 2020-2022	Diferencia	Invarianza
Profesor	0.79	0.84	0.05	Sí
Oral	0.76	0.81	0.05	Sí
Informe	0.72	0.78	0.06	Sí
Empresa	0.85	0.89	0.04	Sí
Auto evaluación	0.68	0.73	0.05	Sí

Todas las diferencias son inferiores a 0.2, confirmando invarianza factorial entre períodos [36]. Esto indica que la estructura del instrumento se ha mantenido estable a lo largo del tiempo.

K. Análisis de tendencia temporal

TABLA XVII
ANÁLISIS DE TENDENCIA TEMPORAL

Modelo	r (año-nota)	p	¿Tendencia?
Tradicional	0.03	0.87	No
Innovado	0.02	0.91	No

La ausencia de tendencia temporal significativa confirma que las diferencias observadas entre modelos son atribuibles al cambio en el instrumento de evaluación, no a efectos de época.

L. Dendograma por cluster

El análisis de componentes por conglomerado mostró diferencias relevantes en la contribución relativa de cada dimensión evaluativa. La evaluación de la empresa y la defensa oral presentaron la mayor variabilidad entre grupos, indicando que estos componentes contribuyen significativamente a la diferenciación de perfiles de desempeño. En contraste, la autoevaluación presentó menor dispersión, sugiriendo una percepción relativamente homogénea del propio desempeño entre los estudiantes.

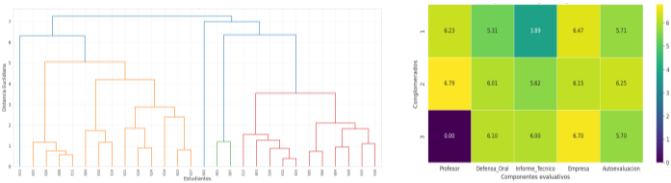


Fig. 11. a) Dendrograma de conglomerados jerárquicos de estudiantes bajo el modelo de evaluación de prácticas basado en SCT. b) Puntajes promedio de los componentes de competencias según conglomerado.

Desde una perspectiva educativa, los patrones observados son consistentes con los principios de evaluación por competencias promovidos por SCT-Chile y los lineamientos internacionales asociados a la formación en ingeniería basada en resultados de aprendizaje. Los resultados sugieren que el modelo rediseñado permite representar de manera más integral las competencias profesionales desarrolladas durante la

práctica, en comparación con enfoques tradicionales centrados únicamente en una calificación global.

L. Tabla resumen completa

TABLA XVIII
RESUMEN DE RESULTADOS

Indicador	Valor	Indicador	Valor
Muestra total	56 estudiantes	Mejora discriminativa	+47.8%
Media Tradicional	6.62	VIF promedio	1.18
Media Innovado	5.84	Correlación Empresa-Nota	0.68
d de Cohen	1.80	Correlación Oral-Nota	0.57
Interpretación d	MUY GRANDE	Correlación Informe-Nota	0.59
CV Tradicional	5.6%	Correlación Profesor-Nota	0.08
CV Innovado	8.3%	Correlación Autoevaluación -Nota	0.36
		Agrupaciones exploratorias	6 (ARI = 0.094)*
		Mejor escenario ABET	Técnico puro (CV 9.6%)

Agrupaciones con baja estabilidad, requieren validación con muestras más amplias.

Los indicadores sintetizados confirman que el modelo innovado presentó mayor capacidad discriminativa, estructura interna consistente y asociaciones coherentes con el desempeño profesional observado. En conjunto, los resultados respaldan la superioridad técnica del enfoque basado en competencias frente al modelo tradicional.

V. DISCUSIÓN

A. Interpretación de los principales hallazgos

El modelo innovado demuestra una capacidad discriminativa significativamente superior, con un coeficiente de variación de 8.3% frente al 5.6% del modelo tradicional, representando una mejora del 47.8%. Este resultado es crucial desde una perspectiva pedagógica, pues indica que el modelo logra diferenciar mejor entre estudiantes de distinto desempeño, permitiendo identificar con mayor precisión fortalezas y áreas de mejora.

La correlación de 0.68 entre la evaluación de la empresa y la nota final valida la importancia de la percepción del empleador en la formación práctica. Este hallazgo se alinea directamente con el RAE 2 de ABET (capacidad para aplicar diseño de ingeniería) [15] y respalda la decisión de asignarle la mayor ponderación (40%).

La baja correlación del componente Profesor (r = 0.08) con la nota final, contrastada con su ponderación del 10%, sugiere la necesidad de revisar este peso en futuras versiones del instrumento, pudiendo transferirse parte de este porcentaje a componentes con mayor poder predictivo como Empresa o Autoevaluación. Por otro lado, la autoevaluación presenta una correlación moderada (r = 0.36) con la nota final, superior a la del componente Profesor, lo que recomienda aumentar su

ponderación al 10-15% en futuras implementaciones, reconociendo su valor formativo y su alineación con el RAE 7 de ABET (aprendizaje continuo).

B. Interpretación de los perfiles exploratorios de clusters

La exploración de agrupaciones estudiantiles sugiere posibles perfiles diferenciados, aunque la baja estabilidad de la solución (ARI = 0.094) impide conclusiones definitivas. Destaca como hipótesis a investigar la posible brecha entre desempeño teórico y práctico observada en una pequeña agrupación (n=2, Cluster 5), que muestra evaluaciones docentes altas (profesor +0.85 DE) contrastando con evaluaciones empresariales muy bajas (empresa -3.07 DE). Este patrón, aunque preliminar, apunta a una posible desconexión entre la formación académica y las demandas del mundo laboral que merece ser estudiada con metodologías cualitativas y muestras más amplias.

La presencia de clusters con n=1 y n=2 es aceptable en la literatura exploratoria cuando representan perfiles atípicos pero teóricamente significativos [37], [38]. Sin embargo, la baja estabilidad (ARI = 0.094) indica que estos perfiles no deben ser utilizados para tomar decisiones pedagógicas sin una validación posterior con muestras más amplias.

C. Implicaciones para la práctica educativa

Los resultados confirman que el modelo innovado SCT se alinea efectivamente con los RAE de ABET:

Empresa (r = 0.68) con RAE 2 (aplicar diseño de ingeniería)

Informe (r = 0.59) con RAE 6 (desarrollar experimentación)

Oral (r = 0.57) con RAE 3 (comunicación efectiva)

Autoevaluación (r = 0.36) con RAE 7 (aprendizaje continuo)

El análisis de sensibilidad sugiere que el escenario "Técnico puro" (55% Empresa, 25% Informe, 10% Oral, 5% Profesor, 5% Autoevaluación) maximizaría la capacidad discriminativa (CV = 9.6%, mejora del 16.8%). Sin embargo, esta optimización debe equilibrarse con consideraciones pedagógicas para mantener el desarrollo de habilidades comunicativas (RAE 3 de ABET). Se recomienda una ponderación balanceada que, sin descuidar la formación técnica, mantenga el énfasis en competencias transversales esenciales para la práctica profesional.

D. Limitaciones y líneas de investigación futura

- Limitaciones del estudio:
 - Diseño de cohortes independientes: Al comparar dos períodos diferentes, no se puede descartar completamente el efecto de variables históricas no controladas, aunque el análisis de tendencia temporal sugiere estabilidad.
 - Tamaño muestral moderado (n=56): Particularmente limitante para el análisis factorial confirmatorio y el análisis de clúster, lo que obliga a interpretar estos resultados con cautela.

Clúster con tamaños pequeños ($n=1$, $n=2$): Aunque teóricamente significativos, requieren validación con muestras más amplias.

Ausencia de validación con criterios externos: No se dispone de medidas independientes del desempeño profesional (ej. evaluaciones de empleadores posteriores al egreso).

- Líneas de investigación futura:

Estudio longitudinal de egresados: Seguimiento de la cohorte innovada para evaluar la relación entre las calificaciones en prácticas y el desempeño profesional posterior.

Validación con muestra ampliada ($n \geq 150$): Para confirmar la estructura factorial y la estabilidad de los perfiles estudiantiles.

Análisis de invarianza por género: Evaluar si el modelo funciona de manera equivalente en subgrupos poblacionales.

Estudio cualitativo con empleadores: Para comprender en profundidad qué aspectos del desempeño evalúan y cómo se relacionan con la formación académica.

VI. CONCLUSIONES

El MI-CT presenta una estructura multidimensional confirmada por baja consistencia interna ($\alpha = 0.105$) y alta independencia entre componentes (correlación promedio = 0.053), lo que indica que cada componente aporta información única sobre el desempeño estudiantil.

La comparación entre cohortes muestra un tamaño del efecto MUY GRANDE ($d = 1.80$, $p < 0.0001$), indicando que los modelos miden constructos fundamentalmente diferentes. El MI-CT demuestra una capacidad discriminativa superior (mejora del 47.8%), permitiendo identificar con mayor precisión diferencias en el desempeño estudiantil.

El análisis factorial confirmatorio, a pesar de las limitaciones propias del tamaño muestral ($KMO = 0.43$), permitió identificar que cuatro de los cinco componentes (Profesor, Informe, Empresa y Autoevaluación) presentan cargas claras en factores únicos, mientras que Oral emerge como un constructo transversal. El cumplimiento del criterio de Fornell-Larcker ($AVE = 0.421 > R^2 \text{ máx} = 0.094$) confirma la validez discriminante del modelo, respaldando su estructura multidimensional pese a las restricciones muestrales.

La exploración preliminar de agrupaciones estudiantiles identificó patrones que sugieren hipótesis sobre la existencia de perfiles diferenciados, particularmente una posible brecha teoría-práctica en un subgrupo de estudiantes. Sin embargo, la baja estabilidad de estas agrupaciones ($ARI = 0.094$) limita la generalización de estos hallazgos, que deben ser validados en estudios futuros con muestras más amplias antes de derivar implicaciones pedagógicas concretas.

El componente Empresa presenta la correlación más alta con la nota final ($r = 0.68$), validando su alineación con el RAE 2 de ABET y respaldando su ponderación del 40% como

el elemento más relevante en la evaluación de prácticas profesionales.

El análisis de sensibilidad reveló que el escenario "Técnico puro" (55% Empresa, 25% Informe, 10% Oral, 5% Profesor, 5% Autoevaluación) maximizaría la capacidad discriminativa ($CV = 9.6\%$, mejora del 16.8%). No obstante, se recomienda un enfoque balanceado que mantenga el desarrollo de habilidades comunicativas (RAE 3 de ABET).

En síntesis, el MI-CT representa un avance conceptual significativo respecto al enfoque tradicional de evaluación en prácticas profesionales de Ingeniería en Construcción, proporcionando evidencia robusta para su implementación en contextos geográficos extremos y su alineación con estándares internacionales de acreditación.

DECLARACIÓN SOBRE EL USO DE TECNOLOGÍAS DE IA GENERATIVA Y ASISTIDA

Durante la preparación de este trabajo, los autores utilizaron ChatGPT 4 (OpenAI) y DeepSeek como herramientas de apoyo para: (1) el análisis estadístico y la generación de código en Python, incluyendo la implementación de análisis factorial exploratorio (AFE) y confirmatorio (AFC), cálculos de fiabilidad (α de Cronbach), correlaciones de Pearson y visualización de datos; y (2) la mejora de la legibilidad y el lenguaje del manuscrito. Estas herramientas contribuyeron a optimizar la eficiencia del procesamiento de datos, la generación de figuras profesionales y la claridad de la redacción. Después de utilizar estas tecnologías, los autores revisaron, validaron y editaron minuciosamente tanto el código generado como el contenido del manuscrito, verificando la corrección de todos los cálculos estadísticos, la precisión de los resultados reportados y la coherencia de las interpretaciones. Los autores asumen toda la responsabilidad por la integridad, precisión y originalidad del trabajo publicado.

AGRADECIMIENTOS

Los autores expresan su sincero agradecimiento a la Red Nacional de Investigación de la Docencia en Educación Superior (RENIDES) y a la Dirección General de Docencia de la Universidad de Magallanes (UMAG), Chile, por el financiamiento otorgado a través de su fondo concursable, permitió distinguir esta propuesta como una de las iniciativas ganadoras para el desarrollo de investigación en docencia universitaria durante el año académico 2026.

Asimismo, agradecemos a los empleadores y estudiantes que participaron voluntariamente en las evaluaciones de prácticas profesionales durante el período 2011-2022, cuya colaboración fue esencial para la realización de este estudio.

Extendemos nuestro reconocimiento al Departamento de Ingeniería en Construcción de la UMAG por facilitar el acceso a los registros institucionales. Confiamos en que los resultados de esta investigación constituirán un aporte significativo para la comunidad universitaria.

DISPONIBILIDAD DE DATOS Y CÓDIGO

Los datos anonimizados que respaldan este estudio (registros de 56 estudiantes, 28 del modelo tradicional 2011-2015 y 28 del modelo innovado SCT 2018-2022), junto con los scripts de Python utilizados para los análisis estadísticos y visualizaciones, están disponibles previa solicitud a los autores. Para acceder a ellos, por favor contactar al autor de correspondencia a través del correo electrónico institucional.

REFERENCIAS

- [1] B. Vivar et al., "Adapting to online education through project-based learning in a complex remote zone," in Proceedings of the LACCEI International Multi-Conference for Engineering, Education and Technology, 2021. doi: 10.18687/LACCEI2021.1.1.123
- [2] B. M. Vivar Vivar et al., "Citizen commitment development through undergraduate internships at the UNPADE-Magallanes," in Proceedings of the LACCEI International Multi-Conference for Engineering, Education and Technology, 2024. doi: 10.18687/LACCEI2024.1.1.456
- [3] C. Martínez, P. Rodríguez, and F. Torres, "Desafíos de la formación en ingeniería en zonas extremas: el caso de la Patagonia chilena," *Revista de Ingeniería y Educación*, vol. 15, no. 2, pp. 112-128, 2020.
- [4] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, 7th ed. Boston, MA: Pearson, 2019.
- [5] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297-334, Sep. 1951. doi: 10.1007/BF02310555.
- [6] J. M. Cortina, "What is coefficient alpha? An examination of theory and applications," *Journal of Applied Psychology*, vol. 78, no. 1, pp. 98-104, 1993. doi: 10.1037/0021-9010.78.1.98.
- [7] N. Schmitt, "Uses and abuses of coefficient alpha," *Psychological Assessment*, vol. 8, no. 4, pp. 350-353, 1996. doi: 10.1037/1040-3590.8.4.350.
- [8] K. Sijtsma, "On the use, the misuse, and the very limited usefulness of Cronbach's alpha," *Psychometrika*, vol. 74, no. 1, pp. 107-120, Mar. 2009. doi: 10.1007/s11336-008-9101-0.
- [9] D. L. Streiner, "Starting at the beginning: An introduction to coefficient alpha and internal consistency," *Journal of Personality Assessment*, vol. 80, no. 1, pp. 99-103, Feb. 2003. doi: 10.1207/S15327752JPA8001_18.
- [10] R. P. McDonald, *Test Theory: A Unified Treatment*. Mahwah, NJ: Lawrence Erlbaum Associates, 1999. doi: 10.4324/9781410601087.
- [11] J. L. Ventura-León and T. Caycho-Rodríguez, "El coeficiente Omega: un método alternativo para la estimación de la confiabilidad," *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, vol. 15, no. 1, pp. 625-627, 2017.
- [12] ABET, "Criteria for Accrediting Engineering Programs, 2026-2027," Engineering Accreditation Commission, Baltimore, MD, 2025.
- [13] ABET, "Criteria for Accrediting Engineering Technology Programs, 2026-2027," Engineering Technology Accreditation Commission, Baltimore, MD, 2025.
- [14] A. Karimi, "A Workshop on ABET-EAC Accreditation Evaluation," in *2025 ASEE-GSW Annual Conference*, Arlington, TX, 2025.
- [15] S. J. Ressler, L. Nolen, and W. R. Bergstrom, "Preparing the Future Civil Engineer: Review and Update of the ABET Civil Engineering Program Criteria," in *2021 ASEE Virtual Annual Conference Content Access*, 2021.
- [16] R. A. Alhorani, W. Abu Elhajja, S. M. Bazlamit, and H. S. Ahmad, "ABET accreditation requirements and preparation: Lessons learned from a case study of Civil Engineering Program," *Cogent Engineering*, vol. 8, no. 1, pp. 1995109, Jan. 2021. doi: 10.1080/23311916.2021.1995109.
- [17] P. Horiuchi and J. Sotomayor, "Methodology to measure student outcomes for ABET accreditation: Case study of an electronic engineering program in Peru," in *2020 IEEE World Conference on Engineering Education (EDUNINE)*, Bogota, Colombia, Mar. 2020, pp. 1-5. doi: 10.1109/EDUNINE48860.2020.9149520.
- [18] A. Azzouni and J. Parham-Mocello, "Evaluating the current continuous improvement approach in abet-accredited computing programs," in Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1, Virtual Event, Germany, Jun. 2021, pp. 78-84. doi: 10.1145/3430665.3456338.
- [19] I. M. Alarifi, "Comparative Analysis on Regional (NCAAA) and International (ABET) Accreditation for Mechanical Engineering Program," *Engineering and Technology Open Access Journal*, vol. 3, pp. 119-134, 2021.
- [20] M. S. Youssef, "The Roles of Different Committees in ABET Accreditation Process for Engineering Programs," *International Journal of Engineering Education*, vol. 4, no. 1, 2022.
- [21] F. M. Lord and M. R. Novick, *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley, 1968.
- [22] M. J. Allen and W. M. Yen, *Introduction to Measurement Theory*. Monterey, CA: Brooks/Cole, 1979.
- [23] J. C. Nunnally and I. H. Bernstein, *Psychometric Theory*, 3rd ed. New York: McGraw-Hill, 1994.
- [24] T. P. Hogan, *Pruebas psicológicas: Una introducción práctica*. México, DF: El Manual Moderno, 2004.
- [25] T. Raykov, "Estimation of maximal reliability: a note on a covariance structure modelling approach," *British Journal of Mathematical and Statistical Psychology*, vol. 57, no. 1, pp. 21-27, May 2004. doi: 10.1348/000711004849295.
- [26] C. Fornell and D. F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," *Journal of Marketing Research*, vol. 18, no. 1, pp. 39-50, Feb. 1981. doi: 10.1177/002224378101800104.
- [27] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 8th ed. Andover, UK: Cengage Learning, 2019.
- [28] G. R. Hancock, L. M. Stapleton, and R. O. Mueller, Eds., *The Reviewer's Guide to Quantitative Methods in the Social Sciences*, 2nd ed. New York: Routledge, 2019. doi: 10.4324/9781315755649.
- [29] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988. doi: 10.4324/9780203771587.
- [30] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley, 1990. doi: 10.1002/9780470316801.
- [31] W. Revelle, "Hierarchical cluster analysis and the internal structure of tests," *Multivariate Behavioral Research*, vol. 14, no. 1, pp. 57-74, Jan. 1979. doi: 10.1207/s15327906mbr1401_4.
- [32] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65, Nov. 1987. doi: 10.1016/0377-0427(87)90125-7.
- [33] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224-227, Apr. 1979. doi: 10.1109/TPAMI.1979.4766909.
- [34] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193-218, Dec. 1985. doi: 10.1007/BF01908075.
- [35] H. Carretero-Dios and C. Pérez, "Normas para el desarrollo y revisión de estudios instrumentales," *International Journal of Clinical and Health Psychology*, vol. 7, no. 3, pp. 863-882, 2007.
- [36] R. J. Vandenberg and C. E. Lance, "A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research," *Organizational Research Methods*, vol. 3, no. 1, pp. 4-70, Jan. 2000. doi: 10.1177/109442810031002.
- [37] C. Snider, S. J. Culley, and E. A. Dekoninck, "Analysing creative behaviour in the later stage design process," *Design Studies*, vol. 34, no. 5, pp. 543-574, Sep. 2013. doi: 10.1016/j.destud.2013.03.001.
- [38] C. Snider, E. A. Dekoninck, and S. J. Culley, "The appearance of creative behavior in later stage design processes," *International Journal of Design Creativity and Innovation*, vol. 2, no. 1, pp. 1-19, Jan. 2014. doi: 10.1080/21650349.2013.807076.