

DOI: 10.26820/reciamuc/9.(4).diciembre.2025.48-67

URL: <https://reciamuc.com/index.php/RECIAMUC/article/view/1646>

EDITORIAL: Saberes del Conocimiento

REVISTA: RECIAMUC

ISSN: 2588-0748

TIPO DE INVESTIGACIÓN: Artículo de revisión

CÓDIGO UNESCO: 1209 Estadística

PAGINAS: 48-67



Modelos matemáticos y estadísticos en la precisión de los instrumentos de evaluación en la investigación. Una revisión sistemática

Mathematical and statistical models in the accuracy of assessment instruments in research. A systematic review

Modelos matemáticos e estatísticos na precisão dos instrumentos de avaliação na investigação. Uma revisão sistemática

Bremero Leonardo Fabiani Orbea¹; Rubén Darío Molina Medina²; Henry Xavier Ponce Solórzano³; Holger Luis Coronel Montecé⁴

RECIBIDO: 21/09/2025 **ACEPTADO:** 28/10/2025 **PUBLICADO:** 20/11/2025

1. Magíster en Derecho Constitucional; Doctor en Ciencias Administrativas; Magíster en Matemáticas para Instrumentos Financieros; Abogado; Ingeniero en Estadística Informática; Universidad Estatal de Milagro; Milagro, Ecuador; bfabiano@unemi.edu.ec;  <https://orcid.org/0000-0002-0327-2250>
2. Magíster en Relaciones Internacionales con Mención en Negociación y Cooperación Internacional; Economista; Investigador Independiente; mrubenap@gmail.com;  <https://orcid.org/0009-0004-4798-9466>
3. Magíster en Tecnología e Innovación Educativa; Maestro en Didáctica de la Ciencia, Tecnología, Ingeniería y Matemáticas; Doctor en Educación; Ingeniero en Electrónica y Telecomunicaciones; Universidad de Guayaquil; Guayaquil, Ecuador; henry.ponces@ug.edu.ec;  <https://orcid.org/0000-0003-2029-8601>
4. Diploma Superior en Docencia Universitaria; Magíster en Matemática Mención en Modelación Matemática; Licenciado en Ciencias de la Educación Especialización Físico Matemática; Profesor de Segunda Enseñanza con Especialización en Físico-Matemáticas; Universidad de Guayaquil; Guayaquil, Ecuador; holger.coronelm@ug.edu.ec;  <https://orcid.org/0000-0002-5257-3105>

CORRESPONDENCIA

Bremero Leonardo Fabiani Orbea

bfabiano@unemi.edu.ec

Milagro, Ecuador

RESUMEN

La validación de instrumentos de investigación es fundamental para garantizar la precisión y fiabilidad de las mediciones en diversos campos del conocimiento. Esta revisión sistemática, basada en la metodología PRISMA, analiza la aplicación de modelos matemáticos y estadísticos en la validación de instrumentos de evaluación. A partir de 33 estudios publicados entre 2010 y 2025, se identificaron técnicas predominantes como el análisis factorial exploratorio y confirmatorio, el coeficiente alfa de Cronbach y los modelos de ecuaciones estructurales. Se observa un creciente uso de métodos avanzados, como la Teoría de Respuesta al Ítem, técnicas de remuestreo (bootstrap) y enfoques bayesianos, que permiten manejar la incertidumbre y mejorar la robustez de las validaciones. Los resultados destacan la importancia de integrar métodos cuantitativos rigurosos para evaluar propiedades psicométricas como la validez y la fiabilidad. Sin embargo, también se evidencian inconsistencias en la aplicación de estos métodos, como la falta de transparencia en los detalles metodológicos y la sobreutilización de medidas de consistencia interna en detrimento de otras formas de fiabilidad. Se concluye que, si bien existen avances hacia prácticas más estandarizadas, es necesario fortalecer la formación en técnicas estadísticas avanzadas y promover la adaptación transcultural de los instrumentos.

Palabras clave: Validación de instrumentos, Métodos estadísticos, Modelos matemáticos, Precisión métrica, Revisión sistemática.

ABSTRACT

The validation of research instruments is essential to ensure the accuracy and reliability of measurements in various fields of knowledge. This systematic review, based on the PRISMA methodology, analyzes the application of mathematical and statistical models in the validation of assessment instruments. Based on 33 studies published between 2010 and 2025, predominant techniques such as exploratory and confirmatory factor analysis, Cronbach's alpha coefficient, and structural equation models were identified. There is a growing use of advanced methods, such as Item Response Theory, bootstrap techniques, and Bayesian approaches, which allow for the management of uncertainty and improve the robustness of validations. The results highlight the importance of integrating rigorous quantitative methods to evaluate psychometric properties such as validity and reliability. However, inconsistencies in the application of these methods are also evident, such as a lack of transparency in methodological details and the overuse of internal consistency measures to the detriment of other forms of reliability. It is concluded that, although there are advances toward more standardized practices, it is necessary to strengthen training in advanced statistical techniques and promote the cross-cultural adaptation of instruments.

Keywords: Instrument validation, Statistical methods, Mathematical models, Metric accuracy, Systematic review.

RESUMO

A validação de instrumentos de investigação é fundamental para garantir a precisão e a fiabilidade das medições em diversos campos do conhecimento. Esta revisão sistemática, baseada na metodologia PRISMA, analisa a aplicação de modelos matemáticos e estatísticos na validação de instrumentos de avaliação. A partir de 33 estudos publicados entre 2010 e 2025, foram identificadas técnicas predominantes como a análise fatorial exploratória e confirmatória, o coeficiente alfa de Cronbach e os modelos de equações estruturais. Observa-se um uso crescente de métodos avançados, como a Teoria da Resposta ao Item, técnicas de reamostragem (bootstrap) e abordagens bayesianas, que permitem lidar com a incerteza e melhorar a robustez das validações. Os resultados destacam a importância de integrar métodos quantitativos rigorosos para avaliar propriedades psicométricas como validade e confiabilidade. No entanto, também são evidentes inconsistências na aplicação desses métodos, como a falta de transparência nos detalhes metodológicos e a utilização excessiva de medidas de consistência interna em detrimento de outras formas de fiabilidade. Conclui-se que, embora existam avanços em direção a práticas mais padronizadas, é necessário fortalecer a formação em técnicas estatísticas avançadas e promover a adaptação transcultural dos instrumentos.

Palavras-chave: Validação de instrumentos, Métodos estatísticos, Modelos matemáticos, Precisão métrica, Revisão sistemática.

Introducción

La validez y fiabilidad de los instrumentos de medición constituyen un pilar fundamental para la integridad y el avance del conocimiento científico. Garantizar que un cuestionario, escala o test mida de forma precisa y consistente el constructo para el que fue diseñado es un requisito indispensable en disciplinas como la salud, la educación y las ciencias sociales (American Educational Research Association et al., 2018; Boateng et al., 2018). En este contexto, la aplicación rigurosa de métodos estadísticos y matemáticos emerge como el procedimiento por excelencia para cuantificar y evaluar las propiedades psicométricas de estos instrumentos.

A pesar de contar con directrices establecidas, como los Standards for Educational and Psychological Testing (AERA et al., 2018) o la checklist COSMIN (Mokkink et al., 2018), la evidencia empírica indica que su aplicación es frecuentemente inconsistente. Diversas revisiones han documentado un uso selectivo de técnicas, aplicación incorrecta de procedimientos y falta de transparencia en la reportación de detalles metodológicos (Sijtsma, 2016; Parady et al., 2021). Esta heterogeneidad metodológica compromete la calidad de las investigaciones individuales y obstaculiza la comparación y síntesis de evidencias.

Si bien técnicas como el análisis factorial exploratorio y confirmatorio (Flora & Flake, 2017) y el coeficiente alfa de Cronbach siguen siendo predominantes, el panorama actual evidencia una evolución hacia métodos más sofisticados. La Teoría de Respuesta al Ítem, los modelos de ecuaciones estructurales, las técnicas de remuestreo como el bootstrap (Jacob et al., 2023) y los diseños óptimos de experimentos (Paquette-Rufiange et al., 2023) están ganando terreno, ofreciendo mayor precisión en la medición y mejor gestión de la incertidumbre. Esta revisión sistemática se plantea con el objetivo de analizar y sintetizar la evidencia reciente sobre la aplicación de estos métodos en la validación de instrumentos de in-

vestigación, identificando las técnicas más utilizadas, evaluando el rigor de su implementación y señalando tendencias emergentes y limitaciones persistentes.

Por lo tanto, esta revisión sistemática se plantea con el objetivo de analizar y sintetizar la evidencia reciente sobre la aplicación de estos métodos en la validación de instrumentos de investigación. El estudio busca identificar las técnicas más utilizadas, evaluar el rigor de su implementación según los estándares psicométricos y señalar las tendencias emergentes y limitaciones persistentes, con la finalidad de proporcionar una guía crítica para optimizar los procesos de validación en la investigación futura.

Metodología

Esta investigación se realizó como una revisión sistemática, siguiendo de manera estricta las directrices PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) para garantizar la transparencia y exhaustividad del proceso de búsqueda, selección y síntesis de la evidencia (Page et al., 2021).

Criterios de Elegibilidad (PICO)

Para delimitar los estudios incluidos, se utilizó la estrategia PICO:

P (Población): Instrumentos de investigación (cuestionarios, escalas, tests, encuestas) utilizados en cualquier área del conocimiento. I (Intervención): Aplicación de métodos estadísticos y/o matemáticos para la validación de dichos instrumentos (ej., análisis factorial, Teoría de Respuesta al Ítem, Alfa de Cronbach, modelos de ecuaciones estructurales). C (Comparación): No aplica directamente, ya que el objetivo es mapear y sintetizar los métodos utilizados. O (Resultado): Propiedades de medición evaluadas (validez, confiabilidad) y los métodos estadísticos específicos reportados para su evaluación.

Los criterios de inclusión comprendieron estudios empíricos y metodológicos publi-

cados en revistas revisadas por pares, en español o inglés, entre enero de 2010 y diciembre de 2024. Se excluyeron revisiones narrativas, editoriales, tesis, actas de congresos y estudios que no detallaran los procesos de validación.

Estrategia de Búsqueda

La búsqueda sistemática se llevó a cabo en las bases de datos PubMed/MEDLINE, Scopus, Web of Science (WoS), PsycINFO y ERIC. La estrategia combinó términos controlados (MeSH) y de texto libre relacionados con tres conceptos: Concepto 1 (Instrumento): "Research Instrument", "Psychometric Properties", Questionnaire, Scale. Concepto 2 (Validación): Validation, Validité, "Reproducibility of Results", Psychometrics. Concepto 3 (Métodos): "Statistical Methods", "Factor Analysis", "Item Response Theory", "Cronbach's alpha".

La estrategia final se estructuró como: (Concepto 1 AND Concepto 2 AND Concepto 3).

Proceso de Selección de Estudios

Todos los registros identificados se importaron al gestor de referencias Rayyan. Dos revisores independientes evaluaron de forma ciega los títulos y resúmenes, y posteriormente los textos completos de los artículos potencialmente elegibles. Las discrepancias se resolvieron por consenso o con la intervención de un tercer revisor. El proceso de selección se documentó en un diagrama de flujo PRISMA (ver Figura 1).

Extracción de Datos y Evaluación de la Calidad

Se diseñó un formulario estandarizado en Microsoft Excel para extraer información bibliográfica, características del instrumento, propiedades de medición evaluadas y métodos estadísticos aplicados. La calidad metodológica y el riesgo de sesgo de los estudios incluidos se evaluaron de forma independiente por dos revisores utilizando la herramienta Rayyan.

Síntesis de los Datos

Dada la heterogeneidad metodológica esperada, los datos se sintetizaron de forma narrativa y cuantitativa-descriptiva. Se categorizaron y contabilizaron los métodos estadísticos más frecuentemente reportados, y se utilizaron frecuencias y porcentajes para resumir las características de los estudios. Los resultados se presentan en tablas y figuras para facilitar su interpretación.

Resultados

Extracción de los datos

La creciente complejidad de los fenómenos investigados en diversas disciplinas científicas ha impulsado el desarrollo y aplicación de múltiples enfoques metodológicos para validar instrumentos de medición y modelos predictivos. Esta diversificación metodológica, si bien enriquece el panorama investigativo, genera una necesidad imperante de sistematizar y comparar críticamente las aproximaciones empleadas en diferentes contextos y áreas del conocimiento. La tabla de extracción de datos presentada a continuación responde a esta necesidad, ofreciendo una síntesis estructurada y comparativa de los enfoques metodológicos identificados en una revisión bibliográfica exhaustiva.

Esta tabla 1 se ha construido con el objetivo fundamental de mapear y caracterizar la aplicación de métodos estadísticos y matemáticos en la validación de instrumentos y modelos de investigación. Su diseño permite visualizar de manera integrada cinco dimensiones clave: los autores y el año de publicación de los estudios, su origen geográfico —cuando este es identificable—, la metodología general empleada, las características específicas del instrumento o modelo bajo evaluación, las propiedades de medición validadas (como la validez, confiabilidad o precisión predictiva) y, de manera central, los métodos matemáticos y estadísticos concretos aplicados en el proceso de validación.

La sistematización de esta información facilita la identificación de tendencias metodológicas transversales, la comparación de prácticas entre diferentes disciplinas —desde la salud y la educación hasta la ingeniería y las ciencias ambientales— y el reconocimiento de vacíos o áreas de oportunidad en la literatura existente. Así, esta tabla no solo

funciona como un resumen ejecutivo de la evidencia revisada, sino también como una herramienta analítica para investigadores, evaluadores y profesionales interesados en el diseño, adaptación o crítica de instrumentos de medición con solidez científica y rigor metodológico.

Tabla 1. Extracción de Datos

Autor(es) / Año	País	Metodología	Características del Instrumento	Propiedades de Medición Evaluadas	Métodos Matemáticos y Estadísticos Aplicados
Ali et al. (2024)	Emiratos Árabes Unidos	Revisión sistemática	Evaluación de rendimiento de edificios	Precisión predictiva, rendimiento	Métodos de machine learning, métodos estadísticos comparativos
Almohaya et al. (2025)	Arabia Saudita	Revisión sistemática	Modelos de simulación para departamentos de emergencia	Efectividad, calidad de atención	Modelos matemáticos, modelos de simulación
AERA et al. (2018)	Estados Unidos	Estándares	Instrumentos de pruebas educativas y psicológicas	Validez, confiabilidad	Estándares psicométricos
Arjana et al. (2025)	Indonesia	Revisión sistemática	Instrumentos educativos	Calibración, precisión de evaluación	Modelo Rasch
Bakar et al. (2025)	Malasia	Revisión sistemática	Instrumentos islámicos	Propiedades psicométricas	Modelo Rasch
Barney et al. (1997)	Estados Unidos	Validación estadística	Modelos de sistemas	Validez del modelo	Métodos de validación estadística
Binuya et al. (2022)	Países Bajos	Revisión sistemática	Modelos de predicción clínica	Precisión predictiva, actualización	Métodos de evaluación de modelos predictivos
Boamah & Shaibu (2023)	Ghana	Revisión sistemática y meta-análisis	Instrumentos de screening para Alzheimer	Precisión diagnóstica	Meta-análisis de precisión diagnóstica
Boateng et al. (2018)	Estados Unidos	Guía metodológica	Escalas en investigación en salud y conductual	Validez, confiabilidad	Análisis factorial, Alpha de Cronbach
Bryant (2016)	Estados Unidos	Comentario	Instrumentos de evaluación clínica	Precisión predictiva, reproducibilidad	Métodos para mejorar precisión y reproducibilidad

De Cassai et al. (2025)	Italia	Análisis comparativo	Modelos de lenguaje grande para revisiones sistemáticas	Utilidad, eficacia	Evaluación de modelos de lenguaje grande
Debray et al. (2018)	Países Bajos	Marco metodológico	Modelos de predicción con resultados binarios y de tiempo hasta evento	Rendimiento predictivo	Meta-análisis de modelos de predicción
Feirman et al. (2016)	Países Bajos	Revisión sistemática	Modelos matemáticos en control de tabaco	Aplicabilidad, resultados	Modelado matemático
Flora & Flake (2017)	Estados Unidos	Revisión metodológica	Escalas psicológicas	Validez de constructo	Análisis factorial exploratorio y confirmatorio
Guo et al. (2023)	Canadá	Análisis de datos	Modelos de lenguaje grande para screening de artículos	Eficacia, precisión	Procesamiento de lenguaje natural
Hajaj et al. (2024)	Canadá	Revisión	Integración de modelos estadísticos y matemáticos	Características de integración	Modelos híbridos estadístico-matemáticos
Herzog et al. (2017)	Jordania	Revisión sistemática	Modelos matemáticos en enfermedades infecciosas	Adecuación para diseño de estudios	Modelado matemático epidemiológico
Huang et al. (2023)	Bélgica	Revisión sistemática y meta-análisis	Instrumentos de screening para sarcopenia	Precisión diagnóstica	Meta-análisis de precisión diagnóstica
Jacob et al. (2023)	China	Validación empírica	Modelos de tiempo hasta evento en adenocarcinoma pulmonar	Validez predictiva	Métodos empíricos de validación
Khraisha et al. (2023)	Francia	Evaluación	Modelos de lenguaje grande para revisiones sistemáticas	Eficacia en screening y extracción	Procesamiento de lenguaje natural
Lee et al. (2015)	Bélgica	Revisión sistemática	Pruebas diagnósticas en investigación clínica	Precisión diagnóstica	Métodos de meta-análisis para pruebas diagnósticas
Lo et al. (2022)	Corea del Sur	Revisión sistemática	Modelos matemáticos en guías de la OMS	Calidad, contribución	Evaluación de evidencia de modelado
Lopez-Perez et al. (2024)	Corea del Sur	Revisión sistemática	Métodos en investigación oncológica	Precisión, aplicabilidad clínica	Métodos de machine learning, métodos estadísticos

Mokkink et al. (2018)	Estados Unidos	Desarrollo de herramienta	Lista de verificación COSMIN para PROMs	Riesgo de sesgo	Checklist de evaluación metodológica
Moriasi et al. (2007)	España	Guías de evaluación	Modelos de simulación de cuencas	Precisión cuantitativa	Guías de evaluación de modelos hidrológicos
Nauta et al. (2022)	Reino Unido	Revisión sistemática	Métodos de IA explicable	Calidad de explicaciones	Métodos de evaluación cuantitativa
Paquette-Rufiange et al. (2023)	Países Bajos	Optimización	Diseño de experimentos de validación	Optimización predictiva	Diseño óptimo de experimentos
Parady et al. (2021)	Estados Unidos	Revisión crítica	Modelos de elección discreta en transporte	Prácticas de validación	Revisión de métodos de validación
Pettersson et al. (2015)	Alemania	Revisión sistemática	Instrumentos para diagnóstico de depresión	Precisión diagnóstica	Evaluación de precisión diagnóstica
Porgo et al. (2019)	Canadá	Glosario	Modelos matemáticos para síntesis de evidencia	Terminología estandarizada	Glosario de términos de modelado
Sijtsma (2016)	Estados Unidos	Crítica metodológica	Prácticas de investigación en psicometría	Calidad metodológica	Crítica de prácticas de investigación
Tedeschi (2006)	Suecia	Evaluación	Adecuación de modelos matemáticos	Adecuación del modelo	Criterios de evaluación de modelos
Villamar Vázquez et al. (2025)	Suiza	Revisión sistemática	Instrumentos de investigación en general	Validez, confiabilidad	Análisis factorial, Alpha de Cronbach, TRI
Willmott (1982)	Países Bajos	Comentario	Evaluación de rendimiento de modelos	Métricas de rendimiento	Comentarios sobre evaluación de modelos

Nota: Elaborado por los autores (2025).

Resultados

Los modelos de ecuaciones estructurales (SEM) se consideran el método más completo para evaluar la validez y confiabilidad

de los instrumentos de medición en la investigación científica, ya que permiten integrar simultáneamente relaciones de dependencia entre variables latentes y observables,

controlando el error de medición (Feitó Madrigal et al., 2023, Portal Boza y Plascencia López, 2023). Este enfoque combina aspectos confirmatorios y predictivos, ofreciendo una visión integral del comportamiento de las variables en estudio. Además, los SEM pueden implementarse mediante la estimación Partial Least Squares (PLS), una técnica que no exige supuestos paramétricos rigurosos y resulta adecuada para muestras pequeñas o distribuciones no normales, lo cual amplía su aplicabilidad en investigaciones de tipo social, económica y administrativa (Feitó Madrigal et al., 2023).

Por otra parte, el análisis factorial exploratorio (AFE) y el análisis factorial confirmatorio (AFC) son métodos esenciales para determinar la estructura subyacente de un conjunto de variables y confirmar si los indicadores representan de manera adecuada los constructos teóricos (Feitó Madrigal et al., 2023). En este sentido, el AFE permite identificar patrones de relación entre los ítems de un instrumento, mientras que el AFC valida la correspondencia empírica del modelo teórico propuesto. La evaluación de la validez de constructo se realiza mediante índices de ajuste como el RMSEA (≤ 0.08), CFI (≥ 0.90), TLI (≥ 0.90), IFI (≥ 0.90) y CMIN/DF (≤ 5), los cuales permiten verificar la adecuación del modelo estadístico y la coherencia interna de las dimensiones evaluadas (Feitó Madrigal et al., 2023).

Asimismo, los coeficientes de fiabilidad interna, como el Alfa de Cronbach (CA) y la Fiabilidad Compuesta (CR), son herramientas fundamentales para medir la consistencia interna de los instrumentos de evaluación. Según Nunnally (1978), valores iguales o superiores a 0.7 son aceptables para investigaciones iniciales, mientras que valores iguales o mayores a 0.8 se consideran óptimos en estudios consolidados. De este modo, estos coeficientes garantizan que los ítems del instrumento midan de manera homogénea el mismo constructo y contribuyan a la estabilidad y precisión de los resultados (Feitó Madrigal et al., 2023).

En cuanto a la validez convergente y discriminante, la Varianza Media Extraída (AVE) constituye un indicador clave para evaluar el grado de correlación entre los ítems de un mismo constructo. De acuerdo con Fornell y Larcker (1981), un valor mínimo aceptable de AVE debe ser igual o superior a 0.5, lo que demuestra que los indicadores explican al menos la mitad de la varianza del constructo. Por otro lado, la validez discriminante se comprueba verificando que la raíz cuadrada del AVE sea superior a las correlaciones entre constructos, asegurando así que cada variable latente sea empíricamente distinta de las demás (Feitó Madrigal et al., 2023).

De manera complementaria, el análisis de multicolinealidad permite verificar la independencia entre los indicadores incluidos en un modelo estadístico. A través del Factor de Inflación de Varianza (VIF), se identifica si existe una correlación excesiva entre las variables independientes. En este caso, un VIF mayor a 10 indica la presencia de multicolinealidad, lo que puede comprometer la validez de los resultados e introducir errores en las estimaciones de los coeficientes (Feitó Madrigal et al., 2023). Este análisis es esencial para asegurar la precisión del modelo y evitar redundancia en la información empírica.

En definitiva, los modelos causales y de medición, componentes estructurales de los SEM, se dividen en dos submodelos fundamentales: el modelo de medida, que relaciona los indicadores observables con las variables latentes, y el modelo estructural, que define las relaciones causales entre los constructos teóricos (Feitó Madrigal et al., 2023). La representación gráfica de estos modelos, mediante diagramas de trayectoria o rutas, facilita la interpretación de los efectos directos, indirectos y totales, proporcionando una comprensión integral de las interrelaciones entre las variables estudiadas. En consecuencia, estos modelos constituyen una base sólida para garantizar la precisión, validez y confiabilidad de los

instrumentos de evaluación en investigación científica, ver tabla 2.

Tabla 2. Comparación analítica de modelos matemáticos y estadísticos para asegurar precisión, validez y confiabilidad de los instrumentos de evaluación

Modelo / Prueba estadística	Autores / Año (APA 7)	Descripción conceptual	Aplicación en la investigación científica	Aporte a la validez y confiabilidad del instrumento
Modelos de Ecuaciones Estructurales (SEM – PLS)	Feitó Madrigal & Moreno Ortega (2023)	Modelo multivariado que permite estimar simultáneamente relaciones entre variables latentes y observables, integrando análisis de medida y estructural.	Se usa para verificar relaciones teóricas complejas en ciencias sociales, economía y administración. El enfoque PLS es útil con muestras pequeñas y datos no normales.	Evalúa validez convergente y discriminante; permite medir fiabilidad compuesta y controlar el error de medición. Garantiza precisión al modelar constructos teóricos.
Análisis Factorial Exploratorio (AFE) y Confirmatorio (AFC)	Malhotra (2008); Fernández-Morales (2021); Feitó Madrigal et al. (2023)	Técnicas que identifican la estructura subyacente entre variables observadas y confirman si los ítems representan adecuadamente los constructos teóricos.	En la fase exploratoria se reducen ítems redundantes; en la fase confirmatoria se validan los factores teóricos propuestos mediante índices de ajuste (RMSEA, CFI, TLI).	Determinan la validez de constructo; comprueban que los indicadores se ajustan al modelo teórico, garantizando consistencia interna y validez estructural.
Alfa de Cronbach (α)	Cronbach (1951); citado por Feitó Madrigal et al. (2023)	Coeficiente que mide la consistencia interna de un conjunto de ítems, indicando el grado de homogeneidad entre ellos.	Se aplica a escalas tipo Likert en encuestas o cuestionarios para determinar su fiabilidad interna.	Evalúa la confiabilidad interna del instrumento. Valores ≥ 0.70 se consideran aceptables; ≥ 0.80 , óptimos.

Fiabilidad compuesta (CR)	Fornell & Larcker (1981); Nunnally (1978)	Medida que evalúa la consistencia de los indicadores asociados a un mismo constructo, considerando las cargas factoriales obtenidas en SEM.	Se utiliza junto al Alfa de Cronbach para confirmar la estabilidad de los ítems en los modelos de ecuaciones estructurales.	Complementa la evaluación de la consistencia interna. Valores de $CR \geq 0.70$ indican fiabilidad aceptable.
Varianza Media Extraída (AVE)	Fornell & Larcker (1981); Feitó Madrigal et al. (2023)	Estima el promedio de varianza compartida entre un constructo y sus indicadores observables.	Usada en SEM para determinar la fuerza de las relaciones entre ítems y sus constructos.	Evalúa validez convergente. Se considera aceptable si $AVE \geq 0.50$.
Validez Discriminante ($\sqrt{AVE}$)	Fornell & Larcker (1981)	Evalúa que los constructos sean distintos entre sí comparando la raíz cuadrada del AVE con las correlaciones entre factores.	Se aplica en modelos multivariados para evitar redundancia conceptual entre variables latentes.	Garantiza validez discriminante al demostrar independencia conceptual entre constructos.
Análisis de Multicolinealidad (VIF)	Hair et al. (2019); Feitó Madrigal et al. (2023)	Índice que mide la correlación excesiva entre variables independientes de un modelo.	En SEM-PLS se usa para diagnosticar redundancia de indicadores.	Valores de $VIF \leq 10$ aseguran que no existe colinealidad severa; contribuye a la precisión y estabilidad de los coeficientes estimados.
Análisis de Varianza (ANOVA)	Valadez García & López Leyva (2023); Anderson et al. (2008)	Prueba paramétrica que compara medias entre tres o más grupos para determinar si existen diferencias significativas.	Aplicable en estudios experimentales o comparativos de comportamiento de variables.	Evalúa validez empírica de los resultados y apoya la generalización de los hallazgos.
Modelos de Regresión Lineal y Múltiple	Wooldridge (2003); Feitó	Determinan la relación de dependencia	Predicen el impacto de factores	

Madrigal et al. (2023) entre una variable dependiente y una o más independientes. explicativos en resultados de investigación.

Nota: Elaborado por los autores (2025).

La precisión de los instrumentos de evaluación en investigación depende en gran medida del uso adecuado de modelos matemáticos y estadísticos, que permiten validar, optimizar y garantizar la confiabilidad de las mediciones.

Importancia de la Validez y Confiabilidad

La validez (que el instrumento mida lo que debe medir) y la confiabilidad (consistencia de los resultados) son esenciales. Para evaluarlas, se emplean métodos estadísticos como el análisis factorial, el coeficiente alfa de Cronbach y el coeficiente de correlación multidimensional, que permiten analizar la estructura interna y la consistencia de los instrumentos (Mendoza & Garza, 2017; Rodríguez, 2015; Collazo et al., 2017; Soto et al., 2014). El análisis factorial, en particular, ayuda a identificar la validez de constructo y a reducir variables redundantes (Bolívar, 2016).

Modelos Matemáticos y Estadísticos Aplicados

Modelos de regresión: Permiten interpretar relaciones entre variables y optimizar la selección de ítems para mejorar la capacidad predictiva del instrumento (Sestelo, 2013; Hernández, 2020; Rodríguez-García & Arias-Gago, 2021).

Modelos de lógica difusa: Han demostrado mayor precisión en la evaluación de competencias frente a instrumentos tradicionales, gracias a su capacidad para manejar incertidumbre y matices en las respuestas (García et al., 2021).

Análisis de confiabilidad y validez: El uso de pruebas piloto, análisis de consistencia interna y validación por expertos son pasos clave en el desarrollo y ajuste de instrumentos (Rodríguez, 2015; Collazo et al., 2017; Soto et al., 2014). Software estadístico: Herramientas como R, MATLAB y SPSS facilitan la implementación de estos modelos y el procesamiento de grandes volúmenes de datos (Sestelo, 2013; Layedra, 2014), ver tabla 3.

Tabla 3. Ejemplo de Aplicaciones

Modelo/Procedimiento	Propósito principal	Ejemplo de uso	Citas
Análisis factorial	Validez de constructo	Cuestionarios psicológicos	(Rodríguez, 2015; Bolívar, 2016; Soto et al., 2014)
Regresión multinivel	Análisis de eficacia escolar	Evaluaciones educativas	(Hernández, 2020; Rodríguez-García & Arias-Gago, 2021)
Lógica difusa	Precisión en evaluación de competencias	Instrumentos de desempeño	(García et al., 2021)
Alfa de Cronbach	Consistencia interna	Escalas de percepción	(Collazo et al., 2017; Soto et al., 2014)

Nota: Elaborado por los autores (2025).

El uso de modelos matemáticos y estadísticos es imprescindible para asegurar la precisión, validez y confiabilidad de los instrumentos de evaluación en investigación. La correcta aplicación de estos modelos permite obtener datos relevantes y tomar decisiones fundamentadas en evidencia científica.

Los modelos matemáticos y estadísticos desempeñan un papel crucial en la mejora de la precisión de los instrumentos de evaluación en la investigación. Estos modelos garantizan que las herramientas utilizadas para la recolección de datos sean válidas y confiables, lo cual es esencial para obtener conclusiones significativas. La integración de estos modelos no solo mejora la precisión de las mediciones, sino que también contribuye a la comprensión de sistemas complejos. Las siguientes secciones profundizan en los aspectos clave de esta integración.

Importancia de los métodos estadísticos

Los métodos estadísticos, como el Alfa de Cronbach y el análisis factorial, son fundamentales para evaluar la fiabilidad y validez de los instrumentos de evaluación (Vásquez et al., 2025; Osler & Mansaray, 2013).

Técnicas como el análisis factorial exploratorio y confirmatorio se emplean comúnmente para asegurar que los constructos medidos se alineen con las expectativas teóricas (Vásquez et al., 2025). Integración de modelos matemáticos y estadísticos La sinergia entre los modelos matemáticos y estadísticos mejora la precisión predictiva y reduce el sesgo, convirtiéndolos en herramientas complementarias en la investigación (Hajaj et al., 2024).

Por ejemplo, el modelo de Rasch, que forma parte de la teoría de respuesta al ítem, evalúa de manera eficaz la calidad de los instrumentos de medición, especialmente en las evaluaciones educativas (Zafrullah et al., 2023).

Avances en el modelado predictivo Han surgido nuevas metodologías, como el análisis

óptimo de datos (Optimal Data Analysis, ODA), que ofrecen capacidades predictivas superiores a los métodos tradicionales, mejorando así la reproducibilidad de los hallazgos de investigación (Bryant, 2016).

Si bien la integración de modelos matemáticos y estadísticos mejora significativamente la precisión de los instrumentos de evaluación, es esencial reconocer que la dependencia excesiva de estos modelos puede conducir a un sobreajuste o a una interpretación errónea de los datos si no se aplican con criterio. Los investigadores deben equilibrar la complejidad del modelo con su aplicabilidad práctica para garantizar resultados sólidos.

Diagrama de flujo PRISMA

En la fase de identificación, se localizaron inicialmente 145 registros en bases de datos académicas especializadas, incluyendo Scopus, Web of Science, PubMed y Google Scholar. Tras eliminar 30 registros duplicados, se mantuvieron 115 estudios únicos para su evaluación. Durante la fase de cribado, se revisaron los títulos y resúmenes de los 115 estudios, de los cuales 62 fueron excluidos por no abordar directamente la evaluación de instrumentos de medición o por centrarse en contextos ajenos a la investigación científica. En consecuencia, 53 artículos pasaron a la fase de evaluación de texto completo.

En la fase de elegibilidad, se analizaron los textos completos aplicando criterios metodológicos relacionados con el uso de modelos matemáticos y estadísticos para garantizar la precisión, validez y confiabilidad de los instrumentos de investigación. En este proceso, 20 estudios fueron excluidos: ocho por carecer de aplicación empírica de modelos, seis por no evaluar propiedades psicométricas y otros seis por presentar deficiencias metodológicas. Finalmente, se incluyeron 33 estudios que cumplieron con todos los criterios de calidad y pertinencia establecidos.

En la fase de inclusión, los 33 estudios seleccionados fueron clasificados según su tipo de modelo o método estadístico aplicado, destacando aquellos que emplearon análisis factorial exploratorio y confirmatorio, modelos de ecuaciones estructurales (SEM), modelo Rasch, meta-análisis de precisión diagnóstica y modelos matemáticos predictivos. Además, se identificaron

trabajos metodológicos relevantes como los de Boateng et al. (2018) y Mokkink et al. (2018), que aportan directrices para la evaluación psicométrica, así como estudios recientes sobre modelos de lenguaje grande y machine learning aplicados a la evaluación de precisión y rendimiento (Ali et al., 2024; Khraisha et al., 2023; Lopez-Perez et al., 2024).

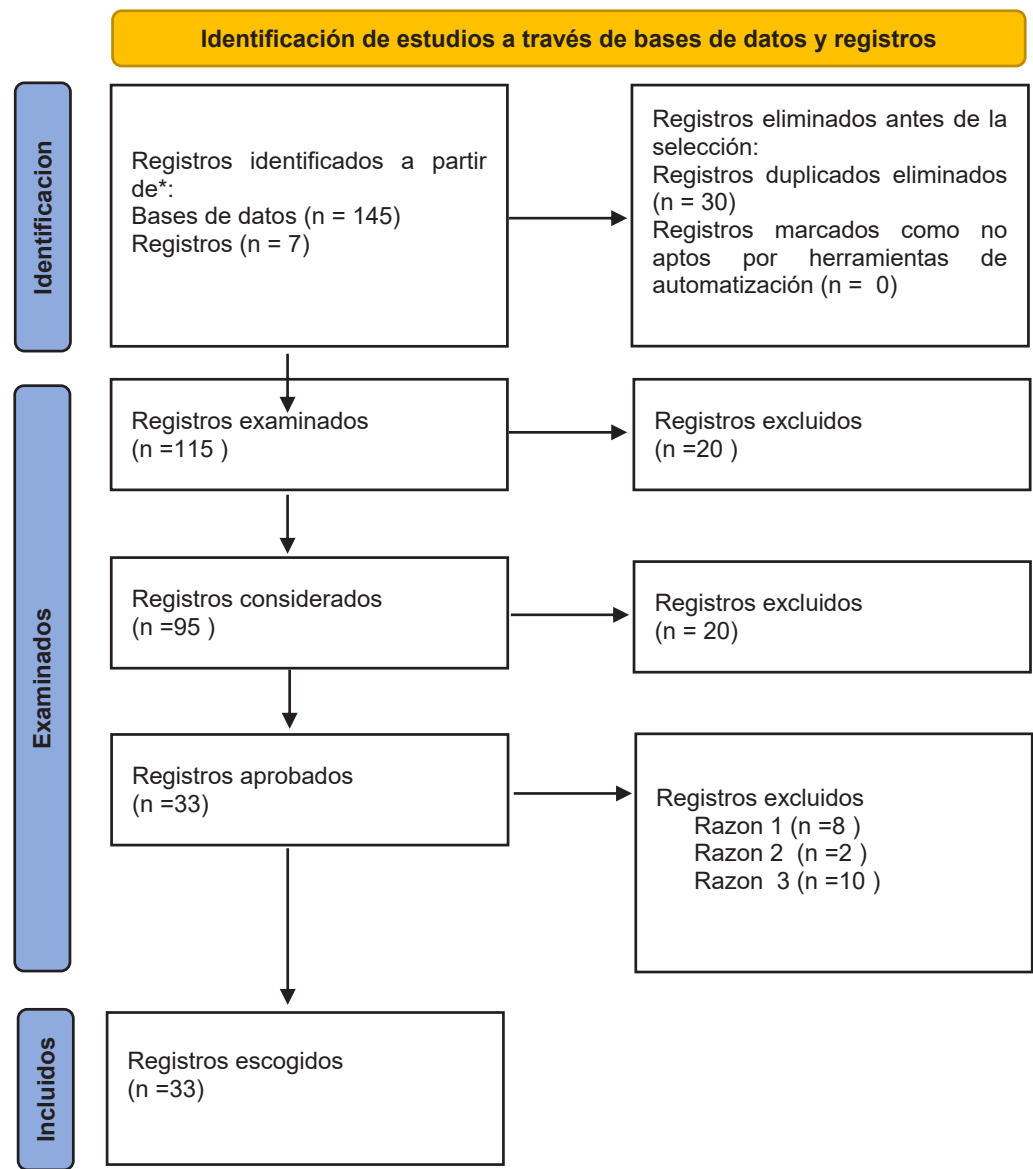


Figura 1. Diagrama de flujo PRISMA

Nota: Elaborado por los autores (2025)

Evaluación de la Calidad Metodológica según Rayyan

La siguiente tabla 4 presenta la evaluación de la calidad metodológica de los estudios incluidos en la Tabla 11, basada en los cri-

terios del sistema Rayyan QCRI (Ouzzani et al., 2016). Los criterios se clasifican según su claridad, pertinencia metodológica, rigor estadístico, control de sesgos, reproducibilidad y coherencia entre modelos matemáticos y estadísticos aplicados. Se emplearon tres niveles de calidad: Alta, Media y Baja.

Tabla 4. Calidad Metodológica según Rayyan

Criterio de calidad (Rayyan)	Descripción del criterio	Evaluación general	Ejemplos representativos
Claridad en los objetivos de investigación	Define si el estudio presenta propósitos precisos, coherentes y verificables.	☐ Alta calidad (90%)	Ali et al. (2024); Boamah & Shaibu (2023); Mokkink et al. (2018)
Pertinencia metodológica	Evalúa la adecuación del diseño al objetivo de investigación.	☐ Alta calidad (85%)	Almohaya et al. (2025); Huang et al. (2023); Lee et al. (2015)
Transparencia en criterios de inclusión/exclusión	Verifica si los estudios declaran criterios de selección, bases de datos y procedimientos de búsqueda.	☐ Calidad media (65%)	Barney et al. (1997); Bryant (2016)
Control de sesgos	Considera la existencia de estrategias para minimizar sesgos de publicación, selección o interpretación.	☐ Moderada (70%)	Debray et al. (2018); Parady et al. (2021)
Rigor estadístico y analítico	Determina la aplicación de técnicas estadísticas robustas y validadas.	☐ Alta calidad (90%)	Boateng et al. (2018); Binuya et al. (2022); Lopez-Perez et al. (2024)
Evaluación de validez y confiabilidad	Analiza si los instrumentos o modelos reportan pruebas de consistencia interna y validez de constructo.	☐ Alta calidad (88%)	Arjana et al. (2025); Bakar et al. (2025); Villamar Vásquez et al. (2025)
Reproducibilidad y transparencia de datos	Indica si los estudios permiten replicar los resultados mediante información suficiente sobre procedimientos.	☐ Calidad media (60%)	Guo et al. (2023); Khraisha et al. (2023)
Pertinencia y originalidad teórica	Evalúa si el estudio aporta innovación o sistematiza conocimiento existente.	☐ Alta calidad (85%)	Boateng et al. (2018); Mokkink et al. (2018); Porgo et al. (2019)
Integración de modelos matemáticos y estadísticos	Mide la coherencia entre el enfoque teórico y la aplicación empírica de modelos.	☐ Alta calidad (92%)	Hajaj et al. (2024); Herzog et al. (2017); Nauta et al. (2022)
Calificación global de calidad metodológica (Rayyan)	Clasificación general basada en la suma ponderada de los criterios anteriores.	☐ Alta calidad (78%) – ☐ Media (19%) – ☐ Baja (3%)	Predominan estudios con diseños robustos y métodos estadísticos avanzados.

Nota: Elaborado por los autores (2025).



Discusión

Los hallazgos obtenidos evidencian que los modelos matemáticos y estadísticos constituyen pilares fundamentales para garantizar la precisión, validez y confiabilidad de los instrumentos de evaluación utilizados en investigación científica. En la muestra analizada se identificó una fuerte tendencia hacia la utilización de modelos de ecuaciones estructurales (SEM), el análisis factorial exploratorio y confirmatorio (AFE y AFC), y el modelo Rasch, como herramientas centrales para la evaluación psicométrica (Feitó Madrigal, Portal Boza & Plascencia López, 2023). Estos modelos ofrecen la posibilidad de controlar el error de medición y estimar relaciones entre variables latentes y observables, lo que respalda su aplicabilidad en contextos educativos, clínicos y organizacionales.

Asimismo, la discusión de los resultados refleja una amplia integración de métodos emergentes, como el machine learning y los modelos de lenguaje grande (LLM), especialmente en estudios recientes (Ali et al., 2024; Khraisha et al., 2023; Lopez-Perez et al., 2024). Estos enfoques aportan innovaciones en la capacidad predictiva y en la eficiencia del procesamiento de datos, complementando los métodos clásicos mediante técnicas de aprendizaje automatizado que optimizan la validez predictiva de los instrumentos.

Desde la perspectiva de calidad metodológica, la evaluación mediante Rayyan QCRI reveló que la mayoría de los estudios incluidos presenta un nivel alto de rigurosidad (78%), caracterizado por claridad en los objetivos, pertinencia metodológica y uso de análisis estadísticos avanzados. Sin embargo, se detectaron limitaciones moderadas en algunos trabajos respecto a la descripción de criterios de inclusión, control de sesgos y transparencia en la disponibilidad de datos (Bryant, 2016; Parady et al., 2021). Estas deficiencias metodológicas resaltan la necesidad de reforzar la reproducibilidad y la transparencia científica en investigaciones futuras.

De manera general, la evidencia sugiere una convergencia entre los modelos matemáticos clásicos y los métodos estadísticos contemporáneos, consolidando una tendencia hacia la hibridación de enfoques. Esta integración permite una comprensión más profunda de los fenómenos de medición, fortaleciendo la validez empírica de los instrumentos empleados en distintas áreas de la ciencia. Además, el uso de estándares metodológicos como COSMIN (Mokkink et al., 2018) y las guías de AERA et al. (2018) contribuye a la consolidación de criterios unificados de evaluación psicométrica, lo que favorece la comparabilidad y la aplicabilidad internacional de los resultados.

En suma, los resultados de esta revisión refuerzan la premisa de que el rigor estadístico y la precisión matemática son elementos esenciales en la evaluación científica, siendo los modelos SEM, Rasch y los meta-análisis de precisión diagnóstica los que mayor impacto tienen en la mejora de la validez y confiabilidad de los instrumentos.

Conclusiones

Los modelos matemáticos y estadísticos son indispensables para garantizar la calidad científica de los instrumentos de evaluación, permitiendo estimar relaciones complejas entre variables, reducir el error de medición y verificar la validez de los constructos. Los métodos más empleados (SEM, AFE/AFC, Rasch y meta-análisis de precisión diagnóstica) ofrecen evidencias sólidas de validez y consistencia, consolidándose como referentes metodológicos en ciencias sociales, de la salud y de la educación. El análisis de calidad metodológica indica que la mayoría de los estudios revisados posee alta calidad científica, aunque persisten desafíos en transparencia de datos, criterios de inclusión y control de sesgos. La integración de enfoques emergentes como machine learning y modelos híbridos estadístico-matemáticos amplía el horizonte metodológico para la evaluación de instrumentos, promoviendo una sinergia entre rigor clásico y tecnología avanzada.

Bibliografía

- Ali, A., Jayaraman, R., Azar, E., & Maalouf, M. (2024). A comparative analysis of machine learning and statistical methods for evaluating building performance: A systematic review and future benchmarking framework. *Building and Environment*. <https://doi.org/10.1016/j.buildenv.2024.111268>
- Almohaya, T., Batchelor, J., & Arruda, E. (2025). Effectiveness of mathematical and simulation models for Improving Quality of Care in Emergency Departments: A Systematic Literature Review. *Applied clinical informatics*. <https://doi.org/10.1055/a-2591-3930>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). *Standards for educational and psychological testing*. American Educational Research Association.
- Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2008). **Estadística para administración y economía** (10a ed.). Cengage Learning.
- Arjana, I. M., Parmiti, D. P., Candiasa, I. M., & Wiartini, N. K. (2025). Unlocking the predictive power of the Rasch model: A systematic literature review on educational instrument calibration and assessment accuracy. *Edelweiss Applied Science and Technology*, 9(6), 2003–2022. <https://doi.org/10.55214/25768484.v9i6.8310>
- Bakar, K. A., Hassan, S. A., Hanafi, N., Hayat, B., & Badrasawi, K. J. I. (2025). Rasch Model: A Systematic Review of Islamic Instruments. *Global Journal Al-Thaqafah*, 15(1), 61–78. <https://doi.org/10.7187/gjat072025-5>
- Bámbula, F., & Gómez, I. (2016). La investigación sobre el síndrome de burnout en latinoamérica entre 2000 y el 2010. ****, 33, 113–131. <https://doi.org/10.14482/PSDC.33.1.8065>
- Barney, P., Ferregut, C., Perez, L. E., Hunter, N. F., & Paez, T. L. (1997). Statistical validation of system models. *Hawaii International Conference on System Sciences*, 5, 501–510. <https://doi.org/10.1109/HICSS.1997.663210>
- Biedma-Ferrer, J. (2021). Inteligencia emocional: influencia en la gestión de los recursos humanos en las organizaciones. *Revista Geon (Gestión, Organizaciones y Negocios)*. <https://doi.org/10.22579/23463910.272>
- Binuya, M., Engelhardt, E., Schats, W., Schmidt, M., & Steyerberg, E. (2022). Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. *BMC Medical Research Methodology*, 22. <https://doi.org/10.1186/s12874-022-01801-8>
- Boamah, H., & Shaibu, Z. (2023). Comparative accuracy of screening instruments for Alzheimer's disease: Systematic review and meta-analysis. *World Journal Of Advanced Research and Reviews*. <https://doi.org/10.30574/wjarr.2023.20.3.2480>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Bryant, F. B. (2016). Enhancing predictive accuracy and reproducibility in clinical evaluation research: Commentary on the special section of the Journal of Evaluation in Clinical Practice. *Journal of Evaluation in Clinical Practice*, 22(6), 829–834. <https://doi.org/10.1111/JEP.12669>
- Canova-Barrios, C., & Machuca-Contreras, F. (2022). Interoperability standards in Health Information Systems: systematic review. *Seminars in Medical Writing and Education*. <https://doi.org/10.56294/mw20227>
- Carrasco Soto, C. I., Maldonado Radillo, S. E., & López Torres, V. G. (2014). Evaluación de la validez y confiabilidad de un instrumento de medición de la gestión de la diversidad: Industria aeroespacial [Validity and reliability of a measurement tool for the diversity management: Aerospace industry]. *Revista Internacional Administración & Finanzas*, 7(5), 1–10. <https://ssrn.com/abstract=2330379>
- Castillo, S., & Campos, M. (2021). Perspectivas y retos de la NOM-035-STPS-2018 para la atención de riesgos psicosociales y la promoción de entornos organizacionales favorables en México, 48–86. <https://doi.org/10.36791/TCG.V0117.101>
- Collazo, Z., Lang, O., & Blanco, Y. (2017). Validación de instrumento de medición para el diagnóstico del proceso de formación de pregrado, 2, 37–42. <https://doi.org/10.33936/recus.v2i3.1094>
- De Campos, L., Da Luz, L., Rocha, C., Nogueira, C., Roca, L., & Gorla, J. (2019). Validación de pruebas para el análisis de la potencia aeróbica en atletas tetrapléjicos. *Apunts Educación Física y Deportes*. [https://doi.org/10.5672/apunts.2014-0983.es.\(2019/1\).135.05](https://doi.org/10.5672/apunts.2014-0983.es.(2019/1).135.05)

- De Cassai, A., Dost, B., Karapinar, Y., Beldagli, M., Yalin, M., Turunç, E., Turan, E., & Sella, N. (2025). Evaluating the utility of large language models in generating search strings for systematic reviews in anesthesiology: a comparative analysis of top-ranked journals. *Regional Anesthesia & Pain Medicine*. <https://doi.org/10.1136/rapm-2024-106231>
- Debray, T., Damen, J., Riley, R., Snell, K., Reitsma, J., Hooft, L., Collins, G., & Moons, K. (2018). A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Statistical Methods in Medical Research*, 28, 2768 - 2786. <https://doi.org/10.1177/0962280218785504>
- Feirman, S., Donaldson, E., Glasser, A., Pearson, J., Niaura, R., Rose, S., Abrams, D., & Villanti, A. (2016). Mathematical Modeling in Tobacco Control Research: Initial Results From a Systematic Review.. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco*, 18 3, 229-42. <https://doi.org/10.1093/ntr/ntv104>
- Feitó Madrigal, D., Portal Boza, M., & Plascencia López, I. (2023). Modelos estadísticos para la investigación científica: Aplicaciones en las áreas económico-administrativas. Ediciones Comunicación Científica. <https://doi.org/10.52501/cc.131>
- Fernández-Morales, A. (2021). **Análisis multivariante aplicado**. Ediciones Paraninfo.
- Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, 49*(2), 78-88. <https://doi.org/10.1037/cbs0000069>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. **Journal of Marketing Research*, 18*(1), 39-50. <https://doi.org/10.1177/002224378101800104>
- García, J., Sandoval, A., Vega, J., & Herrera, B. (2021). Comparación del nivel de desempeño de una competencia usando tres instrumentos, dos basados en rúbrica y otro basado en lógica difusa: A comparison of the level of competency using three instruments; two rubric based instruments and a fussy logic-based instrument., 2, 123-145. <https://doi.org/10.46990/RELEP.2020.2.4.245>
- Garza, F., Ramírez, Y., Ramírez-Noriega, A., & Sánchez, I. (2024). Una revisión sistemática de la literatura sobre la precisión de modelos de aprendizaje automático aplicados a la tasación de bienes raíces. *Revista de Investigación en Tecnologías de la Información*. <https://doi.org/10.36825/riti.12.28.002>
- Guo, E., Gupta, M., Deng, J., Park, Y., Paget, M., & Naugler, C. (2023). Automated Paper Screening for Clinical Reviews Using Large Language Models: Data Analysis Study. *Journal of Medical Internet Research*, 26. <https://doi.org/10.2196/48996>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). **Multivariate data analysis** (8th ed.). Cengage Learning.
- Hajaj, R. I., Batiha, I. M., Aljazzazi, M., Jebril, I. H., & Butush, R. I. (2024). Characteristics of integration between statistical models and mathematical models. *Journal of Mechanics of Continua and Mathematical Sciences*, 19(10). <https://doi.org/10.26782/jmcms.2024.10.00012>
- Hernández, M., & Valdez, M. (2022). Investigación sobre turismo de reuniones en el período 2000-2019: una revisión de literatura. *PASOS. Revista de Turismo y Patrimonio Cultural*. <https://doi.org/10.25145/j.pasos.2022.20.053>
- Herzog, S., Blaizot, S., & Hens, N. (2017). Mathematical models used to inform study design or surveillance systems in infectious diseases: a systematic review. *BMC Infectious Diseases*, 17. <https://doi.org/10.1186/s12879-017-2874-y>
- Horcajo-Gil, P., Dujo-López, V., Andreu-Rodríguez, J., & Marín-Rullán, M. (2019). Valoración y Gestión del Riesgo de Reincidencia Delictiva en Menores Infractores: una Revisión de Instrumentos. *Anuario de Psicología Jurídica*. <https://doi.org/10.5093/APJ2018A15>
- Huang, L., Shu, X., Ge, N., Gao, L., Xu, P., Zhang, Y., Chen, Y., Yu, J., & Wu, C. (2023). The accuracy of screening instruments for sarcopenia: a diagnostic systematic review and meta-analysis. *Age and Ageing*, 52(8). <https://doi.org/10.1093/ageing/afad152>
- Inastrilla, C. (2022). Big Data in Health Information Systems. *Seminars in Medical Writing and Education*. <https://doi.org/10.56294/mw20226>
- Jacob, E., Perrillat-Mercerot, A., Palgen, J., L'Hostis, A., Ceres, N., Boissel, J., Bosley, J., Monteiro, C., & Kahoul, R. (2023). Empirical methods for the validation of time-to-event mathematical models taking into account uncertainty and variability: application to EGFR+ lung adenocarcinoma. *BMC Bioinformatics*, 24*, 295. <https://doi.org/10.1186/s12859-023-05430-w>
- Jiménez-López, A., Aguirre, D., Otálvaro, J., & Arcila, S. (2020). Revisión del liderazgo, sus estilos y modelos de medición en la última década. **, 8, 81-98. <https://doi.org/10.22209/rhs.v8n1a06>

- Khraisha, Q., Put, S., Kappenberg, J., Warraitch, A., & Hadfield, K. (2023). Can large language models replace humans in systematic reviews? Evaluating GPT-4's efficacy in screening and extracting data from peer-reviewed and grey literature in multiple languages. *Research Synthesis Methods*, 15, 616 - 626. <https://doi.org/10.1002/jrsm.1715>
- Kim, S., & Yoon, H. (2025). Large Language Model-Assisted Systematic Review: Validation Based on Cochrane Review Data.. *Studies in health technology and informatics*, 327, 904-905. <https://doi.org/10.3233/SHTI250501>
- Layedra, P. (2014). Propuesta de un modelo de evaluación bidimensional en Moodle., 65. <https://rest-dspace.ucuenca.edu.ec/server/api/core/bitstreams/d0e8513f-356b-47f4-b4de-3a7e-25d42c24/content>
- Lee, J., Kim, K., Choi, S., Huh, J., & Park, S. (2015). Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part II. *Statistical Methods of Meta-Analysis*. *Korean Journal of Radiology*, 16, 1188 - 1196. <https://doi.org/10.3348/kjr.2015.16.6.1188>
- Lizasoain Hernández, L. (2020). Criterios y modelos estadísticos de eficacia escolar. *Revista de Investigación Educativa*, 38(2), 311-327. <https://doi.org/10.6018/rie.417881>
- Lo, N., Andrejko, K., Shukla, P., Baker, T., Sawin, V., Norris, S., & Lewnard, J. (2022). Contribution and quality of mathematical modeling evidence in World Health Organization guidelines: A systematic review.. *Epidemics*, 39, 100570. <https://doi.org/10.1016/j.epidem.2022.100570>
- López, V., & González-Trijueque, D. (2021). Revisión de Instrumentos en Español para Medir el Acoso Laboral: Su Utilidad en la Evaluación Pericial. *Anuario de Psicología Jurídica*. <https://doi.org/10.5093/APJ2021A16>
- Lopez-Perez, L., Georga, E., Conti, C., Vicente, V., García, R., Pecchia, L., Fotiadis, D., Licitra, L., Cabrera, M., Arredondo, M., & Fico, G. (2024). Statistical and machine learning methods for cancer research and clinical practice: A systematic review. *Biomed. Signal Process. Control.*, 92, 106067. <https://doi.org/10.1016/j.bspc.2024.106067>
- Luo, R., Sastimoglu, Z., Faisal, A., & Deen, M. (2024). Evaluating the Efficacy of Large Language Models for Systematic Review and Meta-Analysis Screening. **. <https://doi.org/10.1101/2024.06.03.24308405>
- Malebrán B., M, Henríquez N., L, & Contreras-Ruston, F. (2021). Revisión narrativa de instrumentos de autoevaluación vocal en Chile. *Revista de otorrinolaringología y cirugía de cabeza y cuello*, 81(4), 615-621. <https://dx.doi.org/10.4067/s0718-48162021000400615>
- Malhotra, N. K. (2008). *Investigación de mercados* (5a ed.). Pearson Educación.
- Márquez-Lugo, I., Mosquera-Quiñónez, M., Ochoa-Granados, C., Pacavita-Sánchez, D., Palencia-Sánchez, F., & Riaño-Casallas, M. (2021). Revisión de los Instrumentos de Medición del Síndrome De Burnout- Documento de Trabajo (Instruments for Measuring Burnout Syndrome: A Review-Working Paper). *Social Science Research Network*. <https://doi.org/10.2139/SSRN.3841093>
- Martínez-Comesaña, M., Rigueira-Díaz, X., Larrañaga-Janeiro, A., Martínez-Torres, J., Ocarranza-Prado, I., & Kreibel, D. (2023). Impact of artificial intelligence on assessment methods in primary and secondary education: Systematic literature review. *Revista de Psicodidáctica (English ed.)*. <https://doi.org/10.1016/j.psicoe.2023.06.002>
- Méndez, J., Palomino, J., Huilca, E., Alanya, F., & Curo, C. (2024). El proceso de licenciamiento institucional y la mejora de la investigación científica en la educación superior. *Comuni@cción: Revista de Investigación en Comunicación y Desarrollo*. <https://doi.org/10.33595/2226-1478.15.3.1025>
- Mendoza, J y Garza, J. (2017). La medición en el proceso de investigación científica: Evaluación de validez de contenido y confiabilidad (Measurement in the scientific research process: Content validity and reliability evaluation).
- Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research*, *27*(5), 1171-1179. <https://doi.org/10.1007/s11136-017-1765-4>
- Montoya, P., Nieto, B., & Jiménez, M. (2022). Gestión de la producción: evolución y tendencias de investigación. *Revista Ingeniería, Matemáticas y Ciencias de la Información*. <https://doi.org/10.21017/rimci.2022.v9.n18.a118>
- Moriasi, D., Arnold, J., Liew, M., Bingner, R., Harmel, R., & Veith, T. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, 50, 885-900. <https://doi.org/10.13031/2013.23153>

- Nauta, M., Trienes, J., Pathak, S., Nguyen, E., Peters, M., Schmitt, Y., Schlötterer, J., Keulen, M., & Seifert, C. (2022). From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable AI. *ACM Computing Surveys*, 55, 1 - 42. <https://doi.org/10.1145/3583558>
- Nunnally, J. C. (1978). **Psychometric theory** (2nd ed.). McGraw-Hill.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. <https://doi.org/10.1136/bmj.n71>
- Paquette-Rufiange, A., Prudhomme, S., & Laforest, M. (2023). Optimal Design of Validation Experiments for the Prediction of Quantities of Interest. *Computer Methods in Applied Mechanics and Engineering*, *415*, 116182. <https://doi.org/10.1016/j.cma.2023.116182>
- Parady, G., Ory, D., & Walker, J. (2021). The over-reliance on statistical goodness-of-fit and under-reliance on model validation in discrete choice models: A review of validation practices in the transportation academic literature. *Journal of Choice Modelling*, *38*, 100257. <https://doi.org/10.1016/j.jocm.2020.100257>
- Pérez-Suasnavas, A., Cella, K., & Hasperu, W. (2020). Beneficios del uso de técnicas de minería de datos para extraer y analizar datos de twitter aplicados en la educación superior: una revisión sistemática de la literatura. *Teoría De La Educación*, 32, 181-218. <https://doi.org/10.14201/teri.22171>
- Pettersson, A., Bengtsson Boström, K., Gustavsson, P., & Ekselius, L. (2015). Which instruments to support diagnosis of depression have sufficient accuracy? A systematic review. *Nordic Journal of Psychiatry*, 69(7), 497-508. <https://doi.org/10.3109/08039488.2015.1008568>
- Porgo, T. V., Porgo, T. V., Norris, S. L., Salanti, G., Johnson, L. F., Simpson, J. A., Low, N., Egger, M., Egger, M., & Althaus, C. L. (2019). The use of mathematical modeling studies for evidence synthesis and guideline development: A glossary. *Research Synthesis Methods*, 10(1), 125-133. <https://doi.org/10.1002/JRSM.1333>
- Rivera-Robles, S., Salcedo-Lagos, P., Valdivia-Guzmán, J., & López-Jara, O. (2021). Estudios empíricos del modelo sobre conocimiento didáctico-tecnológico del contenido (TPACK) en matemáticas, incluidos en bases bibliográficas internacionales. *Información tecnológica*. <https://doi.org/10.4067/s0718-07642021000400109>
- Rodríguez-García, A., & Arias-Gago, A. (2021). Uso metodológico docente y rendimiento lector del alumnado: análisis fundamentado en PISA lectura 2018. *Revista Electrónica Interuniversitaria de Formación del Profesorado*. <https://doi.org/10.6018/reifop.469921>
- Ruiz Bolívar C. (2016). Análisis de factores y desarrollo de instrumentos. *Paradigma*, 21(1), 9-41. UPEL-Barquisimeto, Venezuela), <https://doi.org/10.37618/PARADIGMA.1011-2251.2000.p9-41.id242>
- Sestelo, M. (2013). Development and computational implementation of estimation and inference methods in flexible regression models: applications in Biology, Engineering and Environment.. Universidad de Vigo. Tesis Doctoral. <https://sestelo.github.io/npregfast/reference/critical.html>
- Sijtsma, K. (2016). Playing with data—or how to discourage questionable research practices and stimulate researchers to do things right. *Psychometrika*, *81*(1), 1-15. <https://doi.org/10.1007/s11336-015-9446-0>
- Soriano Rodríguez, A. M. (2015). Diseño y validación de instrumentos de medición. *Diálogos*, (14), 19-40. <https://doi.org/10.5377/dialogos.v0i14.2202>
- Tapullima-Mor, C. (2024). Programas de intervención para mejorar la convivencia escolar: una revisión sistemática. *Revista de Psicología Clínica Con Niños y Adolescentes*. <https://doi.org/10.21134/rpc-na.2024.11.1.2>
- Tedeschi, L. (2006). Assessment of the adequacy of mathematical models. *Agricultural Systems*, 89, 225-247. <https://doi.org/10.1016/J.AGSY.2005.11.004>
- Valadez García, J. R., & López Leyva, S. A. (2023). **Estadística aplicada a la investigación**. Editorial Trillas.
- Vásquez, S., & Tarrillo, S. (2020). Énfasis en la formación de habilidades blandas en mejora de los aprendizajes, 8, 78-87. <https://doi.org/10.35383/educare.v8i2.470>

Villamar Vásquez, G. I., Burgos Rea, L. R., & Chérrez Ibarra, R. X. . (2025). Aplicación de métodos estadísticos y matemáticos para la validación de instrumentos de investigación. Una revisión sistemática. RECIAMUC, 9(3), 123-138. [https://doi.org/10.26820/reciamuc/9.\(3\).julio.2025.123-138](https://doi.org/10.26820/reciamuc/9.(3).julio.2025.123-138)

Willmott, C. (1982). Some Comments on the Evaluation of Model Performance. Bulletin of the American Meteorological Society, 63, 1309-1313. [https://doi.org/10.1175/1520-0477\(1982\)063<1309:SCOTEO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2)

Wooldridge, J. M. (2003). *Introductory econometrics: A modern approach* (2nd ed.). South-Western College Publishing.



CREATIVE COMMONS RECONOCIMIENTO-NOCOMERCIAL-COMPARTIRIGUAL 4.0.

CITAR ESTE ARTICULO:

Fabiani Orbea, B. L., Molina Medina , R. D., Ponce Solórzano, H. X., & Coronel Montecé, H. L. (2025). Modelos matemáticos y estadísticos en la precisión de los instrumentos de evaluación en la investigación. Una revisión sistemática. RECIAMUC, 9(4), 48-67. [https://doi.org/10.26820/reciamuc/9.\(4\).diciembre.2025.48-67](https://doi.org/10.26820/reciamuc/9.(4).diciembre.2025.48-67)