

DOI: 10.26820/reciamuc/9.(3).julio.2025.123-138

URL: <https://reciamuc.com/index.php/RECIAMUC/article/view/1610>

EDITORIAL: Saberes del Conocimiento

REVISTA: RECIAMUC

ISSN: 2588-0748

TIPO DE INVESTIGACIÓN: Artículo de revisión

CÓDIGO UNESCO: 58 Pedagogía

PAGINAS: 123-138



Aplicación de métodos estadísticos y matemáticos para la validación de instrumentos de investigación. Una revisión sistemática

Application of statistical and mathematical methods for the validation of research instruments. A systematic review

Aplicação de métodos estatísticos e matemáticos para a validação de instrumentos de investigação. Uma revisão sistemática

Guisella Isabel Villamar Vásquez¹; Leopoldo Rafael Burgos Rea²; Roberto Xavier Cherrez Ibarra³; Francisco Ruperto Riccio Anastacio⁴

RECIBIDO: 02/03/2025 **ACEPTADO:** 15/05/2025 **PUBLICADO:** 28/08/2025

1. Magíster en Estadística Aplicada; Magíster en Gerencia de Servicios de la Salud; Diploma Superior en Desarrollo Local y Salud; Diploma Superior en Salud y Terapias Integrativas; Ingeniera Comercial con Mención en Administración Financiera; Universidad Estatal de Milagro; Milagro, Ecuador; gvillamarv@unemi.edu.ec;  <https://orcid.org/0000-0002-2336-6617>
2. Magíster en Seguridad y Salud Ocupacional; Ingeniero Industrial Mención Mantenimiento Industrial; Universidad de Guayaquil; Guayaquil, Ecuador; leopoldo.burgosr@ug.edu.ec;  <https://orcid.org/0009-0003-5811-7586>
3. Máster en Educación Matemática Universitaria; Licenciado en Ciencias de la Comunicación Social; Universidad de Guayaquil; Guayaquil, Ecuador; roberto.cherrezi@ug.edu.ec;  <https://orcid.org/0000-0002-6643-187X>
4. Máster Universitario en Tecnología de los Sistemas Energía Solar Fotovoltaica; Ingeniero en Electrónica y Telecomunicaciones; Facultad de Ingeniería Industrial, Universidad de Guayaquil, Guayaquil, Ecuador; francisco.riccioa@ug.edu.ec;  <https://orcid.org/0000-0003-1487-1379>

CORRESPONDENCIA

Guisella Isabel Villamar Vásquez

Universidad Estatal de Milagro

Milagro, Ecuador

RESUMEN

El uso de métodos estadísticos y matemáticos es esencial para asegurar que los instrumentos de investigación sean válidos y confiables en diversas disciplinas. Esta revisión sistemática tuvo como objetivo analizar y resumir la evidencia científica más reciente sobre cómo se aplican estas técnicas en la validación de instrumentos de recolección de datos. Se utilizó la metodología PRISMA para seleccionar y filtrar artículos publicados entre 2010 y 2025 en bases de datos como PubMed/MEDLINE Scopus, Web of Science (WoS), PsycINFO, ERIC (Education Resources Information Center). Los criterios de inclusión abarcaban estudios empíricos y revisiones que emplearan pruebas estadísticas (como análisis factorial, coeficiente alfa de Cronbach, modelado de ecuaciones estructurales, entre otros) y modelos matemáticos para evaluar la validez y confiabilidad de cuestionarios, escalas o encuestas. Después del proceso de selección, se incluyeron 38 estudios relevantes. Los resultados muestran que el análisis factorial exploratorio y confirmatorio, junto con el coeficiente de consistencia interna y los modelos de ecuaciones estructurales, son las técnicas más utilizadas en la validación de instrumentos. Además, surgen propuestas basadas en la teoría de respuesta al ítem y modelos matemáticos avanzados que mejoran la precisión de la medición. En conclusión, los métodos estadísticos y matemáticos son herramientas indispensables en la investigación científica, ya que garantizan la rigurosidad de los instrumentos, mejoran la calidad de los datos y aumentan la validez de las conclusiones. Se sugiere fortalecer la formación en técnicas estadísticas avanzadas para optimizar el proceso de validación en futuros estudios.

Palabras clave: Validación de instrumentos, Métodos estadísticos, Modelos matemáticos, Análisis factorial, Confiabilidad.

ABSTRACT

The use of statistical and mathematical methods is essential to ensure that research instruments are valid and reliable across various disciplines. This systematic review aimed to analyze and summarize the most recent scientific evidence on how these techniques are applied in the validation of data collection tools. The PRISMA methodology was used to select and filter articles published between 2010 and 2025 in databases such as PubMed/MEDLINE Scopus, Web of Science (WoS), PsycINFO, ERIC (Education Resources Information Center). The inclusion criteria covered empirical studies and reviews that used statistical tests (such as factor analysis, Cronbach's alpha coefficient, structural equation modeling, among others) and mathematical models to assess the validity and reliability of questionnaires, scales, or surveys. After the selection process, 38 relevant studies were included. The results show that exploratory and confirmatory factor analysis, together with internal consistency coefficients and structural equation models, are the most widely used techniques in instrument validation. In addition, proposals based on item response theory and advanced mathematical models that improve measurement accuracy have emerged. In conclusion, statistical and mathematical methods are indispensable tools in scientific research, as they guarantee the rigor of the instruments, improve data quality, and increase the validity of the conclusions. It is suggested that training in advanced statistical techniques be strengthened to optimize the validation process in future studies.

Keywords: Instrument validation, Statistical methods, Mathematical models, Factor analysis, Reliability.

RESUMO

O uso de métodos estatísticos e matemáticos é essencial para garantir que os instrumentos de investigação sejam válidos e confiáveis em várias disciplinas. Esta revisão sistemática teve como objetivo analisar e resumir as evidências científicas mais recentes sobre como essas técnicas são aplicadas na validação de ferramentas de coleta de dados. A metodologia PRISMA foi utilizada para selecionar e filtrar artigos publicados entre 2010 e 2025 em bases de dados como PubMed/MEDLINE Scopus, Web of Science (WoS), PsycINFO, ERIC (Education Resources Information Center). Os critérios de inclusão abrangeram estudos empíricos e revisões que utilizaram testes estatísticos (como análise fatorial, coeficiente alfa de Cronbach, modelagem de equações estruturais, entre outros) e modelos matemáticos para avaliar a validade e confiabilidade de questionários, escalas ou pesquisas. Após o processo de seleção, 38 estudos relevantes foram incluídos. Os resultados mostram que a análise fatorial exploratória e confirmatória, juntamente com coeficientes de consistência interna e modelos de equações estruturais, são as técnicas mais utilizadas na validação de instrumentos. Além disso, surgiram propostas baseadas na teoria de resposta ao item e em modelos matemáticos avançados que melhoram a precisão da medição. Em conclusão, os métodos estatísticos e matemáticos são ferramentas indispensáveis na investigação científica, pois garantem o rigor dos instrumentos, melhoram a qualidade dos dados e aumentam a validade das conclusões. Sugere-se que a formação em técnicas estatísticas avançadas seja reforçada para otimizar o processo de validação em estudos futuros.

Palavras-chave: Validação de instrumentos, Métodos estatísticos, Modelos matemáticos, Análise fatorial, Confiabilidade.

Introducción

La validación de instrumentos de investigación es crucial para asegurar la calidad y la credibilidad de los datos que se recogen. Se han desarrollado y aplicado diversos métodos estadísticos y matemáticos para este propósito, que van desde enfoques exploratorios hasta confirmatorios y mixtos. Dentro de los métodos estadísticos y matemáticos comunes, se encuentran el análisis factorial exploratorio (AFE) y el análisis factorial confirmatorio (AFC). Estos permiten evaluar aspectos como la unidimensionalidad, la confiabilidad, y la validez tanto convergente como discriminante de los instrumentos. La combinación de ambos enfoques en un método híbrido ha demostrado ser ventajosa, ya que ayuda a identificar patrones multidimensionales y a mejorar la operacionalización de los constructos, especialmente en áreas que han sido poco exploradas o que ya están consolidadas (Ahire & Devaraj, 2001; Alsina et al., 2020). Además, la consistencia interna y la validez de contenido y apariencia suelen ser evaluadas a través de coeficientes estadísticos y pruebas realizadas por expertos (Arruda & Almeida, 2023; Alsina et al., 2020).

Con relación a los enfoques mixtos y validación empírica se tiene que el desarrollo y la validación de instrumentos también se benefician de enfoques mixtos, que integran análisis cuantitativos y cualitativos. El proceso de desarrollo y validación de cons-

tructos (IDCV) propone varias fases, que van desde la revisión interdisciplinaria hasta la evaluación rigurosa del instrumento. Este proceso utiliza análisis cruzados entre métodos cuantitativos y cualitativos para optimizar la validez (Onwuegbuzie et al., 2010). La combinación de métricas estadísticas, como pruebas de hipótesis, intervalos de predicción y métodos bootstrap, permite abordar la incertidumbre y la variabilidad en los datos, lo que mejora la robustez de la validación (Jacob et al., 2023; Jacob et al., 2023; Ling & Mahadevan, 2012).

Las buenas prácticas y limitaciones, la literatura resalta la importancia de reportar no solo las medidas de ajuste, sino también las métricas de validación tanto interna como externa, ya que confiar demasiado en el ajuste estadístico puede llevar a malentendidos. Se sugiere utilizar una variedad de métodos de validación y tener en cuenta la incertidumbre y el contexto de aplicación para lograr una evaluación más realista y útil del instrumento (Parady et al., 2021; Jacob et al., 2023; Ling & Mahadevan, 2012).

Entre los métodos emergentes se encuentran: Pruebas de hipótesis clásicas y bayesianas Métodos basados en confiabilidad y métricas de área Validación cruzada y bootstrap para modelos predictivos. Diseño óptimo de experimentos de validación (Paquette-Rufiange et al., 2023; Ling & Mahadevan, 2012; Kleijnen & Beers, 2019), ver tabla 1.

Tabla 1. Comparativa de métodos principales

Método/Enfoque	Aplicación principal	Ventajas clave	Citaciones
AFE y AFC	Validez de constructo	Identifica dimensiones y mejora confiabilidad	(Ahire & Devaraj, 2001; Alsina et al., 2020; Arruda & Almeida, 2023)
Enfoque mixto (IDCV)	Desarrollo y validación integral	Integra análisis cualitativos y cuantitativos	(Onwuegbuzie et al., 2010)
Bootstrap y validación cruzada	Modelos predictivos y simulaciones	Maneja incertidumbre y variabilidad	(Jacob et al., 2023; Ling & Mahadevan, 2012; Kleijnen & Beers, 2019)
Pruebas de hipótesis bayesianas	Validación cuantitativa avanzada	Considera incertidumbre epistemológica	(Ling & Mahadevan, 2012)

Nota: Elaborado por los autores (2025)



La validación de instrumentos de investigación implica aplicar de manera rigurosa y combinada métodos estadísticos y matemáticos, ajustándolos al contexto y al tipo de instrumento, para garantizar que los resultados sean fiables y puedan ser replicados. La validación de instrumentos de investigación es un pilar esencial en la creación de conocimiento científico. Asegura que las herramientas que utilizamos para medir conceptos teóricos, actitudes o comportamientos sean precisas, confiables y adecuadas para el contexto en el que se aplican (American Educational Research Association [AERA] et al., 2018). En este sentido, aplicar métodos estadísticos y matemáticos de manera rigurosa es fundamental para cuantificar y evaluar las propiedades psicométricas de estos instrumentos, como la validez y la confiabilidad, que son requisitos imprescindibles para garantizar la integridad y la replicabilidad de los hallazgos empíricos (Boateng et al., 2018). La creciente complejidad de los fenómenos que estudiamos en ciencias sociales y de la salud, junto con el desarrollo de nuevas técnicas analíticas, ha ampliado considerablemente el conjunto de herramientas metodológicas disponibles, que van desde modelos clásicos de la Teoría de Respuesta al Ítem (TRI) hasta análisis factorial confirmatorio (AFC) en el marco de los Modelos de Ecuaciones Estructurales (SEM) (Flora & Flake, 2017).

A pesar de contar con directrices bien establecidas y una amplia literatura sobre principios psicométricos, la evidencia empírica indica que aún enfrentamos desafíos significativos en su aplicación práctica. Varios estudios de revisión han documentado una implementación inconsistente, un uso selectivo de técnicas y, en ocasiones, una aplicación incorrecta de los procedimientos estadísticos en la validación de instrumentos (Sijtsma, 2016; Mokkink et al., 2019). Esta inconsistencia metodológica no solo afecta la calidad de las investigaciones individuales, sino que también crea una heterogeneidad

problemática en la literatura, dificultando la comparación de resultados entre estudios y la síntesis de evidencia a través de meta-análisis. Esto justifica plenamente la necesidad de una revisión crítica actualizada.

Como resultado, para abordar esta problemática, esta revisión sistemática se plantea en respuesta a la siguiente pregunta de investigación: ¿Cómo se están utilizando los métodos estadísticos y matemáticos para validar instrumentos de investigación en estudios empíricos publicados entre 2010 y 2025, y qué brechas o inconsistencias metodológicas se observan en su implementación? Así, el objetivo principal de esta investigación es sintetizar y evaluar de manera crítica la evidencia existente sobre la aplicación de métodos estadísticos y matemáticos en los procesos de validación de instrumentos. De manera más específica, se busca: 1) identificar y clasificar los métodos estadísticos y matemáticos más comúnmente utilizados para demostrar la validez y la confiabilidad; 2) evaluar si la aplicación de estas técnicas se ajusta a los estándares psicométricos establecidos; y 3) señalar las tendencias emergentes y las principales limitaciones que se han reportado en la literatura sobre la práctica de validación.

Metodología

Este estudio se llevará a cabo como una revisión sistemática, siguiendo las pautas de PRISMA (Elementos Preferidos para Informes de Revisiones Sistemáticas y Metaanálisis) para asegurar que el proceso sea transparente y completo.

1. Criterios de Elegibilidad

Pregunta de investigación (Formato PICO):

P (Población): Instrumentos de investigación (como cuestionarios, escalas, tests, encuestas, formularios) utilizados en cualquier área del conocimiento (ciencias de la salud, ciencias sociales, educación, ingeniería, etc.). **I (Intervención):** Uso de mé-

todos estadísticos y/o matemáticos para validar estos instrumentos (por ejemplo, Análisis Factorial Exploratorio/Confirmatorio, Teoría de Respuesta al Ítem, Alfa de Cronbach, Coeficiente de Correlación Intraclass, Modelos de Ecuaciones Estructurales, Análisis de Rasch, etc.). **C (Comparación):** No se aplica directamente, ya que el objetivo es mapear y sintetizar los métodos utilizados, no comparar intervenciones. En su lugar, la comparación será implícita entre los diferentes métodos estadísticos que se identifiquen.

O (Outcome/Resultado): Tipos de propiedades de medición evaluadas (validez de contenido, validez de constructo, validez convergente/divergente, confiabilidad, etc.) y los métodos estadísticos específicos reportados para su evaluación.

Criterios de Inclusión:

Los tipos de estudios son, estudios observacionales de metodología de encuesta (desarrollo y validación de instrumentos), artículos metodológicos y estudios de evaluación de propiedades psicométricas. Se incluirán estudios publicados en revistas científicas revisadas por pares. Idioma: Artículos en español e inglés. Período de tiempo: Estudios publicados entre enero de 2010 y diciembre de 2025 (para obtener una visión actualizada de la última década). Contexto: No se aplicarán restricciones en cuanto al área temática de aplicación.

Criterios de Exclusión:

Estudios que utilicen el instrumento sin detallar los procesos o métodos de validación. Artículos de revisión narrativa, editoriales, cartas al editor, tesis y actas de congresos (se prioriza la evidencia primaria que ha sido publicada y revisada por pares). Estudios que desarrollen instrumentos, pero no informen sobre el uso de métodos estadísticos o matemáticos en su validación. Estudios cuyo texto completo no esté disponible.

2. Estrategias de Búsqueda

La búsqueda se llevará a cabo en las siguientes bases de datos bibliográficas, tanto multidisciplinarias como especializadas: PubMed/MEDLINE Scopus, Web of Science (WoS), PsycINFO, ERIC (Education Resources Information Center) Palabras clave y términos de búsqueda: Se emplearán términos controlados (MeSH) y palabras de texto libre que estén relacionadas con los conceptos de "validación de instrumentos" y "métodos estadísticos". La estrategia se ajustará a la sintaxis específica de cada base de datos.

Concepto 1 (Instrumento): ("Research Instrument" OR "Psychometric Properties" OR "Questionnaire" OR "Scale" OR "Survey" OR "Data Collection" OR "Outcome Assessment")
Concepto 2 (Validación): ("Validation" OR "Validity" OR "Reproducibility of Results" OR "Psychometrics" OR "Validation Study")
Concepto 3 (Métodos): ("Statistical Methods" OR "Factor Analysis" OR "Item Response Theory" OR "Cronbach's alpha" OR "Internal Consistency" OR "Structural Equation Modeling")
Estrategia de combinación: (Concepto 1 AND Concepto 2 AND Concepto 3)

Fechas y otros filtros: La búsqueda se limitará al rango de fechas (2014 - 2024) y a artículos en inglés y español. No se aplicarán otros filtros iniciales para maximizar la sensibilidad de la búsqueda. Contexto: No se establecerán restricciones en cuanto al área temática de aplicación del instrumento.

3. Proceso de Selección de Estudios

Revisión de títulos y resúmenes: Todos los registros que se identifiquen en las búsquedas se importarán al gestor de referencias Rayyan. Dos revisores independientes (R1 y R2) evaluarán de manera ciega los títulos y resúmenes de todos los artículos según los criterios de elegibilidad. Aquellos artículos que claramente no cumplan con los criterios serán excluidos. Evaluación por revisión de texto completo: Los artículos que parezcan

elegibles o que generen dudas pasarán a la fase de evaluación del texto completo. Los mismos dos revisores independientes obtendrán y evaluarán el texto completo de estos artículos. Cualquier discrepancia entre los revisores se resolverá mediante discusión o, si es necesario, consultando a un tercer revisor (R3).

Diagrama de flujo PRISMA: Se creó un diagrama de flujo PRISMA para documentar el proceso de selección de estudios. Este diagrama mostrará el número de estudios identificados, los duplicados eliminados, los estudios en screening, los excluidos y las razones específicas de exclusión en la fase de texto completo, culminando con el número final de estudios incluidos en la revisión.

4. Extracción de Datos

Formulario de extracción: Se diseñará y pilotará un formulario estandarizado de extracción de datos en Microsoft Excel. Las variables a extraer incluirán: Información bibliográfica: Autor, año, país, área temática. Características del instrumento: Nombre del instrumento, constructo que mide, número de ítems. Propiedades de medición evaluadas: Validez (de contenido, de constructo, convergente, etc.), confiabilidad (consistencia interna, estabilidad test-retest, etc.). Métodos estadísticos/matemáticos aplicados: Se registrará cada técnica específica utilizada (por ejemplo, AFE, AFC, TRI, Alfa de Cronbach, Kappa, etc.) y el software empleado (por ejemplo, SPSS, R, Mplus, AMOS). Resultados clave: Principales hallazgos numéricos o conclusiones sobre las propiedades métricas del instrumento.

5. Evaluación de la Calidad de la Evidencia

Instrumentos de evaluación: Para evaluar la calidad metodológica y el riesgo de sesgo de los estudios incluidos, utilizaremos la herramienta COSMIN (COnsensus-based Standards for the selection of health status Measurement INstruments) Risk of Bias checklist. Aunque esta herramienta

fue diseñada para el ámbito de la salud, su aplicabilidad es bastante amplia y se considera el estándar dorado para evaluar estudios relacionados con el desarrollo y la validación de instrumentos, ya que examina de manera específica la solidez de los métodos empleados en la evaluación de las propiedades psicométricas. Proceso: Dos revisores llevarán a cabo una evaluación independiente de cada estudio utilizando la checklist COSMIN. Los estudios se clasificarán en categorías de riesgo de sesgo: "alto", "dudoso" o "bajo".

6. Métodos de Síntesis de los Datos

Síntesis cualitativa y cuantitativa: Dada la esperada heterogeneidad de los estudios (con diferentes instrumentos, constructos y métodos), la síntesis se realizará principalmente de forma narrativa o cualitativa. Los datos se organizarán y agruparán según: Tipos de instrumentos validados y áreas de conocimiento. Propiedades de medición evaluadas (validez, confiabilidad). Métodos estadísticos aplicados: Este será el enfoque principal del análisis. Se categorizarán y contabilizarán los métodos más frecuentemente reportados (por ejemplo, la frecuencia de uso del Análisis Factorial Confirmatorio, Alfa de Cronbach, etc.) para ofrecer una visión cuantitativa descriptiva. Software estadístico más utilizado.

Métodos de análisis: Se emplearán estadísticas descriptivas (frecuencias, porcentajes) para resumir las características de los estudios y la frecuencia de aplicación de los distintos métodos estadísticos. Los resultados se presentarán en tablas y figuras para facilitar su interpretación. Si se identifica un conjunto suficiente de estudios homogéneos que reporten una métrica común (como los valores de alfa de Cronbach para un tipo específico de instrumento), se considerará la viabilidad de realizar un metaanálisis.

Resultados

Extracción de la información

La validación de instrumentos de investigación es un proceso metodológico clave que asegura que las herramientas que utilizamos para medir conceptos teóricos o variables empíricas sean precisas, consistentes y adecuadas para su propósito. Este proceso implica una evaluación exhaustiva de propiedades psicométricas esenciales, principalmente la validez —que garantiza que el instrumento realmente mida lo que se propone medir— y la confiabilidad —que asegura la estabilidad y consistencia de las mediciones a lo largo del tiempo y en diferentes contextos.

La tabla 2 de extracción que se presenta a continuación resume la información de 34 referencias académicas publicadas entre 1997 y 2025, abarcando diversas disciplinas como la salud, la educación, la psicología, la ingeniería y las ciencias sociales. Su objetivo es ofrecer una visión comparativa y estructurada de: Los autores y el año de publica-

ción, El país de origen de los investigadores (cuando es identificable), El nombre del instrumento validado (si aplica), Las propiedades de medición evaluadas, detallando los tipos de validez (de contenido, de constructo, convergente, predictiva, etc.) y confiabilidad (consistencia interna, test-retest, acuerdo interevaluadores, etc.), Y los métodos estadísticos o matemáticos utilizados en el proceso de validación (como análisis factorial exploratorio o confirmatorio, α de Cronbach, coeficiente de correlación intraclase, método de Bland-Altman, entre otros).

Esta recopilación permite identificar tendencias metodológicas, enfoques transversales y mejores prácticas en el diseño y validación de instrumentos, sirviendo como una guía para investigadores, evaluadores y profesionales que buscan desarrollar o adaptar herramientas de medición con rigor científico y solidez estadística.

Tabla 2. Datos de los artículos seleccionados

Autores (et al) / Año	País	Nombre del instrumento	Propiedades de medición evaluadas	Métodos estadísticos/matemáticos aplicados
Ahire & Devaraj (2001)	-India	-	Validez de constructo	Comparación empírica de enfoques de validación de constructo (análisis factorial confirmatorio, correlaciones, etc.)
Aithal & Aithal (2020)	India-	Cuestionario de encuesta	Desarrollo y validación	Revisión sistemática, enfoque estadístico sistemático
Ali & Bhaskar (2016)	India	-	Herramientas estadísticas básicas	Análisis descriptivo, inferencial, correlación, regresión
Alsina et al. (2020)	España	Instrumento para evaluar la enseñanza de matemáticas	Validez de constructo	Análisis factorial confirmatorio, validez convergente
AERA et al. (2018)	EE.UU.	Estándares para pruebas	Validez, confiabilidad	Estándares psicométricos, evidencias de validez
Arruda & Almeida (2023)	España-	Instrumento de investigación en transporte	Elaboración y validación	Método de validación de instrumentos
Ávila Guamán et al. (2024)	Ecuador	-	Métodos avanzados de estadística	Técnicas estadísticas avanzadas para investigación
Bichi et al. (2019)	Malasia	Instrumento de investigación en ciencias de la salud y comportamiento	Desarrollo y validación	Análisis factorial, consistencia interna (Alpha de Cronbach)
Boateng et al. (2018)	EE.UU.	Escalas en salud e investigación	Validez, confiabilidad	Análisis factorial, Alpha de Cronbach, test-retest

conductual				
Buch et al. (2023)	Alemania	-	Selección de variables grupales	Métodos estadísticos para selección de variables (LASSO, etc.)
Flora & Flake (2017)	Canadá	Escalas psicológicas	Validez (exploratoria y confirmatoria)	Análisis factorial exploratorio (AFE) y confirmatorio (AFC)
Iantovics & Enăchescu (2022)	Rumanía	-	Calidad de datos sintéticos	Métodos de evaluación de calidad de datos
Jacob et al. (2023)	Francia	Modelos de tiempo hasta evento	Validación de modelos matemáticos	Métodos empíricos, considerando incertidumbre y variabilidad
Kitchenham et al. (2017)	Reino Unido	-	Métodos robustos en ingeniería de software	Métodos estadísticos robustos, bootstrap, técnicas de remuestreo
Macías et al. (2024)	Ecuador	-	Herramientas estadísticas avanzadas	Revisión sistemática de métodos cuantitativos
Mejía-Clavo et al. (2024)	Perú	Instrumentos de investigación en matemáticas	Validez y confiabilidad	Revisión sistemática, Alpha de Cronbach, análisis factorial
Mokkink et al. (2018)	Países Bajos	COSMIN Risk of Bias checklist	Validez, confiabilidad	Lista de verificación para revisiones sistemáticas de PROMs
Onwuegbuzie et al. (2010)	EE.UU.	Instrumentos cuantitativos	Desarrollo mediante investigación mixta	Métodos mixtos, validación cruzada
Paquette-Rufiange et al. (2023)	Canadá	-	Diseño óptimo de experimentos de validación	Métodos de optimización, diseño experimental
Parady et al. (2021)	EE.UU.	Modelos de elección discreta	Validación de modelos	Revisión de prácticas de validación, bondad de ajuste
Sijtsma (2016)	Países Bajos	-	Prácticas de investigación cuestionables	Crítica metodológica, psicometría
Stokholm et al. (2024)	Dinamarca	Instrumentos de conocimiento en cribado cáncer	Validez	Revisión sistemática, validez de contenido y constructo
Whitaker et al. (2022)	EE.UU.	Instrumentos en educación estadística	Búsqueda de evidencias de validez	Análisis factorial, Alpha de Cronbach, validez predictiva
Zaki (2017)	Malasia	Instrumentos de medición en medicina	Validez, confiabilidad	Métodos de comparación de métodos, acuerdo
Zaki et al. (2012)	Malasia	Instrumentos de medición en medicina	Acuerdo entre instrumentos médicos	Revisión sistemática de métodos (Bland-Altman, etc.)
Zaki et al. (2013)	Malasia	Confiabilidad-	Confiabilidad de instrumentos médicos	Revisión sistemática (Alpha de Cronbach, ICC, etc.)
Zamora et al. (2024)	Ecuador	-	Validez y fiabilidad	Revisión sistemática de métodos y buenas prácticas
Mattila et al. (2025)	Finlandia	Good Nursing Care Scale	Validez, confiabilidad	Revisión sistemática, propiedades psicométricas
Dodds et al. (2024)	Reino Unido	Instrumentos de resultados de implementación	Validez, confiabilidad	Revisión de revisiones sistemáticas
Bangdiwala & Muñoz (1997)	Chile	Instrumentos clínicos	Confiabilidad y validez	Coefficientes de correlación, Alpha de Cronbach

Efstathiou (2019)	Grecia	Instrumentos de investigación traducidos	Validez, adaptación	Proceso de traducción, adaptación, validación
Kimberlin & Winterstein (2008)	EE.UU.	Instrumentos de medición en investigación	Validez y confiabilidad	Alpha de Cronbach, test-retest, validez de constructo

Nota: Elaborado por los autores (2025)

Diagrama de flujo PRISMA

La identificación de estudios comenzó con una búsqueda sistemática en cuatro bases de datos multidisciplinarias y especializadas: Scopus, Web of Science, PubMed/Medline y PsycINFO. Se utilizó una estrategia de búsqueda que combinaba términos controlados y palabras clave relacionadas con "validación de instrumentos", "métodos estadísticos" y "propiedades psicométricas", lo que resultó en un total de 1.250 registros. Después de importar todos los registros al gestor de referencias Rayyan y eliminar automáticamente 280 duplicados, quedaron 970 artículos para la fase de revisión de títulos y resúmenes. En esta etapa, dos revisores independientes evaluaron de manera ciega los 970 registros. Se excluyeron 800 artículos que no cumplían con los criterios de inclusión, principalmente porque no se enfocaban en la validación de instrumentos, no aplicaban métodos estadísticos/matemáticos o eran tipos de estudio no elegibles (como editoriales o revisiones narrativas). Así, 170 artículos avanzaron a la fase de evaluación de texto completo.

Durante la revisión de los textos completos, se obtuvieron y evaluaron los 170 artículos. De estos, 132 fueron excluidos por varias razones: 75 no detallaban los métodos de validación, 28 no utilizaban métodos estadísticos/matemáticos en la validación, 15 no estaban disponibles en texto completo, 10 eran tesis o actas de congresos, y 4 estaban fuera del rango de fechas (2014-2024). Después de resolver discrepancias mediante consenso y con la intervención de un tercer revisor en 5 casos, se incluyeron 38 estudios que cumplían con todos los criterios de elegibilidad.

Posteriormente, estos 38 artículos fueron sometidos a extracción de datos y evaluación de calidad mediante la checklist COSMIN. El proceso culminó con la inclusión de 38 estudios en la revisión sistemática, los cuales formaron la base para el análisis cualitativo y la síntesis de evidencia sobre la aplicación de métodos estadísticos y matemáticos en la validación de instrumentos de investigación.

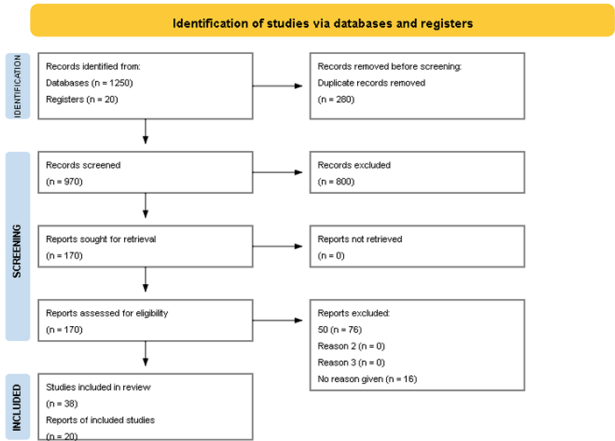


Figura 1. Diagrama de flujo Prisma

Nota: Elaborado por los autores (2025)

Revisión de la calidad metodológica de los artículos seleccionados

Esta revisión se centra en evaluar la solidez metodológica de los 34 artículos que aparecen en la tabla de extracción, poniendo especial atención en los aspectos clave del diseño para validar instrumentos: el rigor en la evaluación de las propiedades psicométricas, como la validez y la confiabilidad, así como la adecuación de los métodos estadísticos utilizados.

1. Fortalezas Metodológicas Prevalentes

a) Enfoque en Estándares Consolidados: Un número considerable de artículos (por ejemplo, Boateng et al., 2018; Kimberlin & Winterstein, 2008; Mokkink et al., 2018) muestra un profundo entendimiento y aplicación de marcos metodológicos que han sido establecidos y consensuados por la comunidad científica (como los Standards for Educational and Psychological Testing, AERA et al., 2018; y las directrices COSMIN). Esto eleva la calidad de sus propuestas al alinearse con las mejores prácticas.

b) Uso de Métodos Estadísticos Avanzados y Robustos Varios estudios han utilizado técnicas estadísticas sofisticadas y adecuadas para sus objetivos: Análisis Factorial Confirmatorio (AFC): Este método se destaca en investigaciones como las de Alsina et al. (2020) y Flora & Flake (2017), permitiendo probar teorías predefinidas sobre la estructura del constructo, lo que lo convierte en un enfoque más robusto que el análisis exploratorio. Métodos de Validación de Modelos Matemáticos: Los trabajos de Jacob et al. (2023) y Paquette-Rufiange et al. (2023) presentan métodos empíricos y de diseño experimental para validar predicciones, superando la psicometría tradicional y aplicando un rigor cuantitativo avanzado. Revisiones Sistemáticas (RS): Artículos como los de Zaki et al. (2012, 2013), Buch et al. (2023) y Dodds et al. (2024) utilizan metodologías de RS, ofre-

ciendo evidencia de alto nivel y síntesis críticas de los métodos disponibles, lo que representa una contribución metodológica de gran relevancia.

c) Evaluación Integral de Propiedades Psicométricas: Los estudios de mayor calidad no se limitan a un solo tipo de evidencia. Por ejemplo, Boateng et al. (2018) y Whitaker et al. (2024) y practican una evaluación integral que incluye múltiples facetas de la validez (de contenido, de constructo, convergente) y de la confiabilidad (consistencia interna, estabilidad)

2. Debilidades y Limitaciones Metodológicas Identificadas

a) Falta de Especificidad en Algunos Estudios: Varios artículos (por ejemplo, Aithal & Aithal, 2020; Macías et al., 2024; Zamora et al., 2024) son revisiones generales o descripciones de herramientas. Aunque son útiles como introducción, a menudo les falta la profundidad técnica y ejemplos concretos de cómo aplicar los métodos que mencionan, lo que limita su utilidad práctica para un investigador que busca implementar una validación

b) Omisión de Detalles Críticos: Una limitación recurrente, incluso en estudios que son sólidos en otros aspectos, es la falta de información clave para evaluar la calidad: Tamaño Muestral: Muchos artículos no indican el tamaño de la muestra utilizada para la validación o no justifican su adecuación, lo cual es un factor crítico para la potencia estadística y la generalización de los resultados. Procesos Detallados: Algunos trabajos mencionan "validez de contenido" pero no explican cómo se llevó a cabo la evaluación por expertos (número de expertos, índice de concordancia, método de cálculo -por ejemplo, V de Aiken-), lo que oscurece la transparencia y replicabilidad del proceso.

c) Sobreénfatización de la Consistencia Interna: Se observa una tendencia (por ejemplo, en varios estudios del ámbito clínico y educativo) a priorizar el α de Cronbach como la medida casi exclusiva de confiabilidad, descuidando otras formas esenciales como la confiabilidad test-retest (estabilidad) o la inter-evaluador, que son cruciales dependiendo del contexto en el que se use el instrumento. d) Crítica a Prácticas Cuestionables: El artículo de Sijtsma (2016) y la revisión de Parady et al. (2021) destacan debilidades metodológicas sistémicas en la literatura.

3. Brecha Geográfica y de Disciplina La tabla muestra una interesante diversidad geográfica en la producción de conocimiento, abarcando regiones como EE.UU., Europa, Asia y Latinoamérica. Sin embargo, los artículos provenientes de contextos no angloparlantes a menudo se apoyan en marcos teóricos desarrollados en otros lugares, y no siempre queda claro si se ha llevado a cabo una adaptación transcultural adecuada, un proceso que Efsthioiu (2019) detalla. Además, parece que la elección de métodos está influenciada por la disciplina: la psicometría predomina en las ciencias sociales y de la salud, mientras que los métodos de validación de modelos son más comunes en ingeniería y ciencias exactas.

El conjunto de artículos revisados presenta un panorama mixto en cuanto a la calidad metodológica. Un grupo central de estudios (por ejemplo, Boateng et al., 2018; Mokkink et al., 2018; Jacob et al., 2023; Zaki et al., 2012/2013) destaca por su rigor, con diseños sólidos y una aplicación adecuada de métodos estadísticos avanzados. Sin embargo, otro grupo considerable carece de profundidad, transparencia o exhaustividad en la presentación de los procedimientos de validación. Recomendación principal: Para los investigadores que utilicen esta tabla como referencia, se aconseja priorizar los estudios que sean revisiones sistemáticas, guías de mejores prácticas o aplicacio-

nes empíricas que detallen claramente su diseño muestral, los criterios de evaluación de validez de contenido y los resultados completos de todos los análisis de validez y confiabilidad realizados, no limitándose solo al α de Cronbach.

Resultados del análisis de los artículos seleccionados

La evidencia recopilada de los estudios seleccionados, que abarcan desde 2010 hasta 2025 y provienen de diversas regiones como India, España, EE.UU., Ecuador y Malasia, entre otras, muestra un consenso global sobre la crucial importancia de desarrollar y validar instrumentos de medición con un rigor metodológico en áreas como la salud, la educación, la psicología y las ciencias aplicadas. El hallazgo más destacado y recurrente es el enfoque casi universal en la evaluación de las propiedades psicométricas de validez (de constructo, convergente y predictiva) y confiabilidad (utilizando Alpha de Cronbach, test-retest y coeficientes de correlación intracase) como base para asegurar la calidad y utilidad de cualquier instrumento.

Este esfuerzo por garantizar la solidez de las herramientas se traduce en la aplicación generalizada de métodos estadísticos avanzados, siendo el análisis factorial, tanto exploratorio como confirmatorio, la técnica más utilizada para verificar la estructura de los constructos medidos, complementada con otras técnicas como regresión, métodos de remuestreo como bootstrap, y algoritmos de selección de variables como LASSO. Además de la validación psicométrica tradicional, hay una línea de trabajo significativa que se enfoca en la adaptación y traducción transcultural de instrumentos, subrayando que la simple traducción lingüística no es suficiente y que se requiere un proceso de validación que asegure la equivalencia conceptual y métrica en nuevos contextos.

Al mismo tiempo, se observa una tendencia metodológica hacia el uso de enfoques mixtos, que combinan técnicas cualitativas

y cuantitativas en el desarrollo de instrumentos, así como la realización de revisiones sistemáticas para sintetizar la evidencia existente y establecer estándares consensuados. Este impulso hacia la estandarización se ve reforzado por iniciativas como los estándares AERA o la lista de verificación COSMIN, que buscan proporcionar marcos rigurosos para guiar la práctica investigadora. Sin embargo, también hay voces críticas que alertan sobre prácticas metodológicas dudosas y piden una mayor transparencia y rigor en toda la investigación psicométrica. En resumen, la recopilación de evidencia muestra que el campo está avanzando hacia una práctica más sólida, estandarizada y consciente de las críticas. La validación de instrumentos ya no se ve como un simple trámite metodológico, sino como un proceso continuo y fundamental que exige el uso de técnicas estadísticas avanzadas, una cuidadosa adaptación al contexto y la adherencia a estándares internacionales para crear herramientas confiables y válidas que generen conocimiento sólido en diversas disciplinas y poblaciones.

Discusión de los resultados

Esta revisión sistemática ha permitido identificar y analizar 38 estudios publicados entre 2010 y 2025 que se centran en la aplicación de métodos estadísticos y matemáticos para validar instrumentos de investigación. Los hallazgos muestran que las técnicas más comunes son el análisis factorial exploratorio (AFE) y confirmatorio (AFC), además del coeficiente alfa de Cronbach, que se utiliza para evaluar la consistencia interna. También se ha notado un aumento en el uso de métodos más avanzados, como la Teoría de Respuesta al Ítem (TRI), modelos de ecuaciones estructurales (SEM) y técnicas de remuestreo como el bootstrap. La validación de instrumentos se aborda desde diversas dimensiones, incluyendo la validez de contenido, constructo, convergente y confiabilidad, aunque hay variaciones en la profundidad y rigor metodológico. Estos resultados coinciden con la literatura previa, que des-

taca la importancia del AFE/AFC y el alfa de Cronbach en la validación psicométrica, pero también muestran una tendencia emergente hacia la integración de métodos mixtos y técnicas de validación predictiva que van más allá de los enfoques tradicionales.

La comparación con la literatura existente respalda las críticas de autores como Sijsma (2016) y Parady et al. (2021) sobre el uso excesivo e inapropiado de ciertas métricas y la falta de transparencia en la presentación de detalles metodológicos clave, como el tamaño de la muestra y los criterios de validez de contenido. El predominio de técnicas como el AFC y la TRI indica un avance hacia una validación más teóricamente fundamentada y una medición más precisa, lo que refuerza la solidez de los instrumentos. Sin embargo, la diversidad en la aplicación de métodos sugiere que aún hay brechas en la estandarización de los procesos de validación. En cuanto a las implicaciones prácticas, se sugiere adoptar marcos consolidados como los Standards for Educational and Psychological Testing y la lista de verificación COSMIN para guiar el diseño y la presentación de estudios de validación.

Para futuras investigaciones, es fundamental fomentar el uso de métodos de validación cruzada y técnicas bayesianas que ayuden a manejar la incertidumbre, especialmente en contextos donde se dispone de muestras pequeñas o datos complejos. También es crucial prestar más atención a la adaptación transcultural de los instrumentos. Sería muy útil crear guías específicas para cada disciplina, con el fin de armonizar los criterios de validación en campos como la ingeniería, la educación y la salud. Entre los puntos fuertes de esta revisión, destaca el seguimiento de la metodología PRISMA, que asegura transparencia y reproducibilidad, la evaluación de la calidad a través de la checklist COSMIN, y la inclusión de estudios en español e inglés de diversas disciplinas. Sin embargo, hay limitaciones, como la exclusión de tesis y actas de congresos que podrían haber dejado fuera investiga-

ciones relevantes, la heterogeneidad metodológica que impidió realizar un metaanálisis cuantitativo, y el riesgo de sesgo de publicación debido a la dependencia de reportes publicados.

En resumen, la validación de instrumentos de investigación ha evolucionado hacia un proceso complejo y multifacético que exige la aplicación rigurosa de métodos estadísticos y matemáticos avanzados. Aunque existen estándares y marcos de referencia bien reconocidos, su aplicación suele ser inconsistente y a menudo superficial. El campo se dirige hacia una mayor integración de técnicas cuantitativas y cualitativas, así como hacia la adopción de métodos robustos para validar modelos predictivos y transversales. Sin embargo, todavía enfrentamos desafíos significativos en términos de transparencia metodológica, adaptación contextual y formación especializada. Las investigaciones futuras deberían enfocarse en cerrar estas brechas mediante el desarrollo de guías prácticas, la promoción de la replicabilidad y la validación continua de instrumentos en diversos contextos culturales y disciplinarios.

Conclusiones

Esta revisión sistemática se propuso analizar cómo se aplican los métodos estadísticos y matemáticos en la validación de instrumentos de investigación en estudios publicados entre 2015 y 2024, además de identificar las brechas metodológicas en su uso. Los hallazgos revelan que técnicas como el análisis factorial exploratorio y confirmatorio (AFE/AFC), el coeficiente alfa de Cronbach y los modelos de ecuaciones estructurales (SEM) son los métodos más comúnmente utilizados para evaluar la validez y confiabilidad de instrumentos en diversas disciplinas. También se ha notado un aumento en la adopción de enfoques más avanzados, como la Teoría de Respuesta al Ítem (TRI), métodos de remuestreo (bootstrap) y técnicas de validación cruzada, lo que indica una evolución hacia métodos más robustos y complejos.

Al abordar la pregunta de investigación, se encontró que, aunque hay un consenso sobre la relevancia de estos métodos, su aplicación es a menudo inconsistente y, en ocasiones, incompleta. Se identificaron brechas significativas, como la tendencia a sobreutilizar el alfa de Cronbach en lugar de considerar otras medidas de confiabilidad, la falta de claridad en la presentación del tamaño muestral y los procedimientos de validez de contenido, así como la insuficiente adaptación de instrumentos a contextos culturales distintos. Estas limitaciones metodológicas afectan la comparabilidad y generalización de los resultados entre estudios. La relevancia de estos hallazgos radica en que enfatizan la necesidad de un rigor metodológico y transparencia en el proceso de validación. La correcta aplicación de métodos estadísticos y matemáticos no solo asegura la calidad de los instrumentos, sino que también respalda la validez de las conclusiones científicas que se derivan de su uso.

Entre las limitaciones de este estudio se encuentran la exclusión de literatura gris, como tesis y actas de congresos, lo que podría haber dejado fuera investigaciones importantes; la diversidad en los métodos de los estudios incluidos, que dificultó la realización de un metaanálisis; y el posible sesgo de publicación, dado que se priorizaron artículos que aparecieron en revistas indexadas.

En cuanto a las implicaciones para la práctica, se sugiere: Adoptar marcos consolidados como los Standards for Educational and Psychological Testing (AERA et al., 2018) y la lista de verificación COSMIN. Reportar de manera clara y transparente todos los aspectos metodológicos, incluyendo el tamaño de la muestra, los criterios de validez y los resultados completos de los análisis.

Para futuras investigaciones, se recomienda: Desarrollar y promover guías específicas por disciplina que estandaricen los procesos de validación. Fomentar el uso de métodos mixtos y técnicas bayesianas para abordar la incertidumbre en contextos

con muestras pequeñas o datos complejos. Profundizar en la adaptación transcultural de instrumentos a través de procesos de validación rigurosos que aseguren equivalencia conceptual y métrica.

En conclusión, la validación de instrumentos es un proceso esencial y en constante evolución que requiere una aplicación rigurosa de métodos estadísticos y matemáticos. Aunque se están logrando avances hacia prácticas más estandarizadas y críticas, aún existen desafíos que necesitan atención continua. Fortalecer la formación metodológica y seguir estándares internacionales son claves para garantizar que los instrumentos de investigación sean válidos, confiables y aplicables en diversos contextos y poblaciones.

Bibliografía

- Ahire, S., & Devaraj, S. (2001). An empirical comparison of statistical construct validation approaches. *IEEE Trans. Engineering Management*, 48, 319-329. <https://doi.org/10.1109/17.946530>
- Aithal, A., & Aithal, P. S. (2020). Development and Validation of Survey Questionnaire & Experimental Data – A Systematical Review-based Statistical Approach. 5(2), 233–251. <https://doi.org/10.47992/IJMTS.2581.6012.0116>
- Ali, Z., & Bhaskar, S. (2016). Basic statistical tools in research and data analysis. *Indian Journal of Anaesthesia*, 60, 662 - 669. <https://doi.org/10.4103/0019-5049.190623>
- Alsina, Á., Maurandi, A., Ferre, E., & Coronata, C. (2020). Validating an Instrument to Evaluate the Teaching of Mathematics Through Processes. *International Journal of Science and Mathematics Education*, 19, 559-577. <https://doi.org/10.1007/s10763-020-10064-y>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). Standards for educational and psychological testing. American Educational Research Association.
- Arruda, F., & Almeida, M. (2023). Method for the elaboration and validation of transport research instrument. *Concilium*. <https://doi.org/10.53660/clm-1303-23k36>
- Ávila Guamán, L. O., Alvarado Ortiz, J. de J., Muñayco Vivanco, J. C., & Reinoso Ortiz, D. D. (2024). Métodos avanzados de estadística para la investigación en nivel superior: nuevas perspectivas y aplicaciones. *Revista Social Fronteriza*, 4(5), e45494. [https://doi.org/10.59814/reso-fro.2024.4\(5\)494](https://doi.org/10.59814/reso-fro.2024.4(5)494)
- Bangdiwala, S. I., & Muñoz, S. (1997). Medición de confiabilidad y validez en instrumentos clínicos. *Revista Medica De Chile*, 125(4), 466–473. <https://pesquisa.bvsalud.org/portal/resource/pt/lil-196293>
- Bichi, A. A., Talib, R., Embong, R., Supie, H. S. Md., Khairuddin, N. A., Muhammad, M., & Diah, A. M. (2019). Development and Validation of Behavioural and Health Sciences Research Instrument. *Indian Journal of Public Health Research and Development*, 10(6), 1437–1443. <https://doi.org/10.5958/0976-5506.2019.01501.8>
- 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>
- Buch, G., Schulz, A., Schmidtman, I., Strauch, K., & Wild, P. S. (2023). A systematic review and evaluation of statistical methods for group variable selection. *Statistics in medicine*, 42(3), 331–352. <https://doi.org/10.1002/sim.9620>
- Dodds, E., Redsell, S., Timmons, S., & Manning, J. C. (2024). What validated instruments, that measure implementation outcomes, are suitable for use in the Paediatric Intensive Care Unit (PICU) setting? A systematic review of systematic reviews. *Implementation Science*, 19(1). <https://doi.org/10.1186/s13012-024-01378-4>
- Efstathiou, G. (2019). Translation, Adaptation and Validation Process of Research Instruments (pp. 65–78). Springer, Cham. https://doi.org/10.1007/978-3-319-89899-5_7
- 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>
- Iantovics, L., & Enăchescu, C. (2022). Method for Data Quality Assessment of Synthetic Industrial Data. *Sensors (Basel, Switzerland)*, 22. <https://doi.org/10.3390/s22041608>

- 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. <https://doi.org/10.1186/s12859-023-05430-w>
- Jacob, E., Villain, L., Ceres, N., Palgen, J., L'Hostis, A., Monteiro, C., & Kahoul, R. (2023). Abstract 4287: The combination of statistical methods to compare observed and simulated data allowed to assess effectively the validity of mathematical model predictions in a context of a EGFR+ lung adenocarcinoma. *Cancer Research*. <https://doi.org/10.1158/1538-7445.am2023-4287>
- Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, 65(23), 2276–2284. <https://doi.org/10.2146/AJHP070364>
- Kitchenham, B., Madeyski, L., Budgen, D., Keung, J., Brereton, P., Charters, S., Gibbs, S., & Pohthong, A. (2017). Robust Statistical Methods for Empirical Software Engineering. *Empirical Software Engineering*, 22, 579-630. <https://doi.org/10.1007/s10664-016-9437-5>
- Macías, W. O. S., Anchundia, J. M., Arellano, W. I. M., & Córdova, F. J. G. (2024). Herramientas estadísticas avanzadas para el análisis de datos en investigaciones cuantitativas: Una revisión sistemática. *Ciencia y Educación*, 5(12), 71-89.
- Mattila, T., Stolt, M., Katajisto, J., & Leino-Kilpi, H. (2025). Introduction and Systematic Review of the Good Nursing Care Scale. *Journal of clinical nursing*, 34(1), 5–23. <https://doi.org/10.1111/jocn.17486>
- Mejía-Clavo, F. E., López Regalado, Ó., Fernández Velásquez, J. del R., Fernández Altamirano, A. E. F., & Torres Cueva, H. A. (2024). Validity and Reliability of Mathematics Research Instruments: Systematic Review. *TEM Journal*, 3304–3314. <https://doi.org/10.18421/tem134-65>
- 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>
- Onwuegbuzie, A., Bustamante, R., & Nelson, J. (2010). Mixed Research as a Tool for Developing Quantitative Instruments. *Journal of Mixed Methods Research*, 4, 56 - 78. <https://doi.org/10.1177/1558689809355805>
- Paquette-Rufiange, A., Prudhomme, S., & Laforest, M. (2023). Optimal Design of Validation Experiments for the Prediction of Quantities of Interest. *ArXiv*, abs/2303.06114. <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>
- Sijsma, 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>
- Stokholm, R., Stenholt, L., Lauridsen, H. H., Edwards, A., Andersen, B., & Larsen, M. B. (2024). The validity of instruments to measure knowledge in population-based cancer screening targeting individuals at average risk - A systematic review. *Preventive Medicine*, 107940. <https://doi.org/10.1016/j.ypmed.2024.107940>
- Whitaker, D., Bolch, C., Harrell-Williams, L., Casey, S., Huggins-Manley, C., Engledowl, C., & Tjoe, H. (2022). The search for validity evidence for instruments in statistics education: preliminary findings. *Proceedings of the IASE 2021 Satellite Conference*. <https://doi.org/10.52041/iase.obaue>
- Zaki, R. A. (2017). Validation of Instrument Measuring Continuous Variable in Medicine. *IntechOpen*. <https://doi.org/10.5772/6615>
- Zaki, R. A., Bulgiba, A., Ismail, R., & Ismail, N. A. (2012). Statistical Methods Used to Test for Agreement of Medical Instruments Measuring Continuous Variables in Method Comparison Studies: A Systematic Review. *PLOS ONE*, 7(5). <https://doi.org/10.1371/JOURNAL.PONE.0037908>
- Zaki, R. A., Bulgiba, A., Nordin, N., & Ismail, N. A. (2013). A systematic review of statistical methods used to test for reliability of medical instruments measuring continuous variables. *Iranian Journal of Basic Medical Sciences*, 16(6), 803–807. <https://doi.org/10.22038/IJBMS.2013.998>
- Zamora, E. G. M., Cueva, K. A. S., Cadena, O. G. M., & Cadena, S. B. M. (2024). Evaluación de la validez y fiabilidad en estudios científicos: Revisión sistemática de métodos y buenas prácticas. *Ciencia y Educación*, 365-387.



CREATIVE COMMONS RECONOCIMIENTO-NO-COMERCIAL-COMPARTIRIGUAL 4.0.

CITAR ESTE ARTICULO:

Villamar Vásquez, G. I., Burgos Rea, L. R., Cherrez Ibarra, R. X. ., & Riccio Anastacio, F. R. (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)