

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), marzo-abril 2026,
Volumen 10, Número 2.

https://doi.org/10.37811/cl_rcm.v10i2

ESTRATEGIAS PARA EL DISEÑO Y VALIDACIÓN DE INSTRUMENTOS: UNA GUÍA PRÁCTICA PARA LA INVESTIGACIÓN INSTRUMENTAL

**STRATEGIES FOR THE DESIGN AND VALIDATION
OF INSTRUMENTS: A PRACTICAL GUIDE FOR
INSTRUMENTAL RESEARCH**

Roberto Joel Tirado-Reyes

Universidad Autónoma de Sinaloa, México

Rosario Eduardo Apodaca-Armenta

Universidad Autónoma de Sinaloa, México

Rosario Edith Ortiz Félix

Universidad Autónoma de Sinaloa, México

Jorge Luis Garcia Sarmiento

Universidad Autónoma de Sinaloa, México

Edgar Fabián Torres Hernández

Universidad de Guanajuato, México

Estrategias para el Diseño y Validación de Instrumentos: Una Guía Práctica para la Investigación Instrumental

Roberto Joel Tirado-Reyes¹

robertojr@uas.edu.mx

<https://orcid.org/0000-0002-1492-7507>

Universidad Autónoma de Sinaloa
México

Rosario Eduardo Apodaca-Armenta

eduardoapodaca@uas.edu.mx

<http://orcid.org/0000-0002-8144-3457>

Universidad Autónoma de Sinaloa
México

Rosario Edith Ortiz Félix

rosarioortiz@uas.edu.mx

<https://orcid.org/0000-0002-5827-3218>

Universidad Autónoma de Sinaloa
México

Jorge Luis García Sarmiento

jorge.garcia@uas.edu.mx

<https://orcid.org/0000-0001-6650-1762>

Universidad Autónoma de Sinaloa
México

Edgar Fabián Torres Hernández

ef.torres@ugto.mx

<https://orcid.org/0000-0001-9596-634X>

Universidad de Guanajuato
México

RESUMEN

La presente investigación propone el uso de estrategias innovadoras en el diseño y la validación de instrumentos de medición, con un enfoque en el área de la salud. Se invoca aquí la integración de métodos cualitativos y cuantitativos, ya que de esta forma se logra que la validez de contenido y constructo se obtenga de forma robusta, además de alta confiabilidad y adaptación transcultural. El objetivo principal es elaborar una plantilla que permita a los investigadores diseñar instrumentos que reduzcan los sesgos y garanticen rigor en los procedimientos de la recolección y el análisis de datos. Así, el estudio propone que se adopten enfoques sistemáticos y de múltiple determinación que incluyan la evaluación de peer review y con métodos estadísticos de alta complejidad para aumentar la reproducibilidad y la validez de los instrumentos, de forma que se actúe de manera significativa a la calidad de los resultados investigativos.

Palabras clave: diseño de instrumentos, validación de instrumentos, psicometría, adaptación transcultural, ciencias de la salud

¹ Autor principal.

Correspondencia: rosarioortiz@uas.edu.mx

Strategies for the Design and Validation of Instruments: A Practical Guide for Instrumental Research

ABSTRACT

This study introduces innovative strategies for the design and validation of measurement instruments within the health sciences. It emphasizes the integration of qualitative and quantitative methodologies, a combination that strengthens both content and construct validity while ensuring high reliability and cultural adaptability. The main objective is to provide a structured framework that enables researchers to design instruments capable of reducing bias and safeguarding methodological rigor in data collection and analysis. The proposal highlights the importance of systematic and multidimensional approaches that incorporate peer review processes alongside advanced statistical methods to increase reproducibility and validity. By adopting these strategies, the research seeks to contribute meaningfully to the enhancement of scientific rigor, offering practical tools that support the development of reliable instruments and, ultimately, improve the quality of research outcomes in the health field.

Keyword: instrument design, instrument validation, psychometrics, transcultural adaptation, health sciences

*Artículo recibido 02 abril 2026
Aceptado para publicación: 30 abril 2026*



INTRODUCCIÓN

El desarrollo y validación de instrumentos de medición son aspectos esenciales dentro de la investigación científica, especialmente en disciplinas como la psicología, educación y ciencias de la salud; dichos instrumentos, que incluyen cuestionarios, escalas e inventarios, permiten la recopilación de datos de manera estructurada y fiable, asegurando una representación precisa de los constructos que se buscan medir (Maldonado-Suárez & Santoyo-Telles 2024; González-Ramírez, Matus-Miranda & Fernández-García 2017). El proceso de validación de un instrumento debe seguir una serie de pasos que progresan jerárquicamente con respecto a consideraciones como la validez de contenido, consistencia y la fiabilidad de sus partes; en cuanto a validación de contenido, se corrobora a través de una evaluación por un panel de expertos que valoran los ítems por su precisión, importancia y qué también representan el constructo en cuestión (Lira & Caballero 2020; Rodríguez et al, 2024). Al mismo tiempo, la validez del constructo se refuerza a través de varios análisis estadísticos, como análisis factorial exploratorio y confirmatorio que permiten evaluar la estructura latente de los datos para confirmar si el instrumento en cuestión mide lo que se supone que debe medir (Romero & Mora 2020; Vallejo-Gómez et al, 2021) Específicamente en el campo de la salud, el diseño del instrumento enfrenta desafíos específicos debido a la multiplicidad de los contextos en los que se utiliza; en este sentido, la adaptación transcultural de instrumentos ha surgido como un enfoque crítico para acomodar el propósito previsto en diferentes grupos poblacionales objetivo; dicha adaptación va más allá de una traducción precisa, también incluye la validación del significado del constructo en diversos contextos, evitando así el sesgo semántico y cultural (Maytorena, Reyes & Sánchez, 2023; Torres-Hernández, 2022). Relacionado con lo anterior la consistente viabilidad, junto con la estabilidad temporal del instrumento, son aspectos que mencionan la confiabilidad, la homogeneidad y correlación entre los ítems pueden ser medidos con el Alfa de Cronbach y el Omega de McDonald (Supo, 2013). Así como también el análisis test-retest, la validez de criterio, apriori y a posteriori, ofrecen pruebas de relevancia al contexto del objeto de estudio (Rodríguez et al, 2024; Romero & Mora 2020).

Es necesario mencionar que la revisión de las normativas disponibles para la validación de instrumentos de medición revela que, a pesar de existir múltiples protocolos, aún quedan importantes deficiencias; por ejemplo el uso de técnicas de medición a nivel interdisciplinario sigue teniendo severas



limitaciones, en la comprensión de los ítems, debido al predominio de los enfoques cuantitativos que en la integración sistemática de métodos cualitativos; por otra parte también se evidencia falta de consenso en la aplicación de criterios rigurosos para la medición de la claridad y representatividad de los reactivos, así como en la aplicación de técnicas factoriales avanzadas que aseguren la replicabilidad de la estructura interna en muestras diversas; además, resulta fundamental la creación de protocolos unificados que en su medición combinen validez predictiva y validez concurrente, en los procesos de adaptación transcultural; es por esto que el enfoque de estas cuestiones pone de manifiesto la necesidad de reforzar y actualizar los métodos de validación para generar instrumentos confiables y funcionales en una variedad de contextos de investigación (Maldonado-Suárez & Santoyo-Telles 2024; Romero & Mora 2020; Torres-Hernández, 2022).

Este estudio busca generar impactos positivos en la calidad de la investigación al ofrecer una guía práctica sobre el diseño de estrategia innovador, así como la validación de instrumentos de medición; así como profundizar en la elaboración de estos instrumentos, analizando la incorporación de métodos cualitativos y cuantitativos necesarios para su validación y fiabilidad; pero sobre todo y no menos importante formular propuestas que faciliten el desarrollo de herramientas de medición para los investigadores que requieran un adecuado estudio de la comunidad científica.

Tipos de Instrumentos

Las herramientas de recopilación de datos se diferencian por su estructura y objetivo. En su clasificación, existen algunas que son muy comunes para el estudio en ciencias de la salud y disciplinas secundarias, tales como: Cuestionarios: en la investigación, su mayor ventaja es la combinación de preguntas cerradas y abiertas, como también su análisis cuantitativo y cualitativo. Esto permite captar muchos matices en la percepción, conocimiento y actitudes de los participantes. En contraste, escalas e inventarios, debido a su diseño rígido, se centran mayormente en los análisis cuantitativos y el uso de los estadísticos (Tsang & Terkawi 2017). Escalas: como la de Likert y las escalas visuales análogas, permiten la medición de las emociones, actitudes, y percepciones en un rango específico, las cuales son llamadas variables latentes (González-Ramírez, & García-Hernández, 2022). Inventarios: estas son listas que permiten a los profesionales valorar ciertos parámetros en el estudio, tales como habilidades, conocimiento, o síntomas clínicos, en el caso de las mujeres (Boparai, & Kathuria, 2018).



Estos instrumentos juegan un papel clave en la enfermería y las ciencias de la salud porque permiten la evaluación de habilidades clínicas, el bienestar del paciente y la calidad de la atención; su correcta construcción y validación son importantes para garantizar que la recopilación de datos sea representativa y confiable.

Relevancia del Diseño y Validación de Instrumentos

El enfoque sistemático hacia la creación y validación de herramientas de medición es particularmente relevante para la investigación científica en salud, educación y ciencias sociales, ya que las herramientas bien estructuradas ayudan a recopilar datos adecuados, precisos y representativos, mientras reducen sesgos y errores sistemáticos que de otro modo amenazarían la validez de los hallazgos (González-Ramírez, Matus-Miranda & Fernández-García 2017; Fenn, Tan & George 2020). En el cuidado de la salud, validar cuestionarios y escalas es fundamental para medir síntomas, calidad de vida relacionada con la salud y adherencia a la terapia para evaluaciones completas que informen decisiones importantes tomadas por los profesionales de la salud (Maldonado-Suárez & Santoyo-Telles 2024).

En las conversaciones en disciplinas de la salud, evaluar competencias clínicas y habilidades interpersonales utilizando escalas o inventarios estructurados es, sin lugar a dudas, central para avanzar en la formación profesional, así como en la calidad en la atención al paciente. Esto acontece porque la falta de instrumentos válidos y fiables podría comprometer la evaluación del desempeño, afectando tanto la docencia como la práctica clínica (Aljehani, et al, 2023; Apodaca-Armenta, Gómez-Rodríguez & Rodríguez-Vázquez 2024). Igualmente, en el ámbito de la salud pública y epidemiología, el empleo de herramientas de medición en el estudio de determinantes sociales, hábitos de vida y la prevalencia de enfermedades en la población aseguran que los hallazgos sean extrapolables y comparativos entre grupos poblacionales (Bautista-Espinel et al, 2022).

La falta de adecuadas pruebas de validación en los instrumentos afecta la calidad de la ciencia, porque si los cuestionarios o escalas no han sido sometidos a pruebas de validación y confiabilidad, los datos obtenidos pueden ser imprecisos o no representativos. Entre los métodos que pueden ayudar a solventar ese problema, es mejor usar como mínimo, el análisis factorial exploratorio y confirmatorio que se propuso para ayudar a la confiabilidad de las herramientas y que el cuestionario realmente mide el constructo de interés (Romero & Mora 2020).



En adición, la investigación es más replicable si los instrumentos han sido diseñados y validados porque se pueden usar en otros estudios sin perder la calidad en las mediciones (Ranganathan 2022).

En el ámbito de la salud, los dispositivos de medición impactan tanto en la política pública como en el quehacer clínico, dado que estos instrumentos permiten medir la evaluación que tiene el paciente sobre su tratamiento, la efectividad de las intervenciones terapéuticas, así como el nivel de conocimientos y competencias de los profesionales de la salud (Al-Madaney, & Fässler, 2023; López-Belmonte, 2022).

La estricta validación de herramientas garantiza que la información que se obtiene y se recoge, especialmente en distintos momentos y lugares, es confiable y por ende robusta y se puede utilizar para el desarrollo de evidencia científica (Fenn, Tan & George 2020). Debido a esto, el diseño y validación de instrumentos son considerados en terapia ocupacional como pilares en la investigación y en la práctica profesional porque permiten que las mediciones sean precisas, y en consecuencia, la calidad metodológica de los estudios se enriquezca aumentando la posibilidad de que los resultados se apliquen en diferentes situaciones clínicas y científicas (Madsø, Pachana, & Nordhus 2023).

Fase Cualitativa: Validez Interna del Instrumento

A medida que se dispone de nuevas fuentes de datos, tratar de asegurarse de que los datos recopilados para el estudio reflejan con precisión la realidad del fenómeno en estudio se vuelve aún más importante. Dentro de esta área general, se incluyen varias estrategias de validación, como la validación de contenido, la validación racional, la validación por jueces y la validación de respuesta. La validación de contenido se enfoca en determinar si los temas se relacionan con el área de conocimiento o el dominio de conducta de interés; para cumplir con esto, se sugiere hacer evaluaciones de literatura, contraponer con instrumentos existentes, y hacer discusiones con expertos en la materia (Boparai, & Kathuria, 2018), uno de los métodos de mayor resonancia para la validación del contenido de un instrumento es la revisión por expertos. En este caso, los especialistas examinan los ítems por su relevancia y certifican que la información es clara y los ítems son representativos (Rodríguez et al, 2024). También es posible aplicar cuantitativamente el coeficiente V de Aiken para estimar el acuerdo de los jueces con la relevancia del ítem y alcanzar un juicio evaluativo cuantitativo en su conjunto (Maldonado-Suárez & Santoyo-Telles 2024). Este último método mejora la objetividad de la decisión respecto a la idoneidad de los ítems incluidos.



El enfoque de validez racional o validez lógica se centra en la lógica y la teoría de la validación. Esta validación se ocupa de todos los ítems y de la lógica interna de los ítems con el constructo medido. Todo ítem debe alinearse con la operación definicional del fenómeno a estudiar (Lira & Caballero 2020). Con el objetivo de mejorar esta forma de validez, se proponen mapas conceptuales o diagramas que relacionen los ítems y las dimensiones del constructo (Ranganathan 2022). También es recomendable que se realice una prueba piloto con una muestra representativa para detectar problemas de incoherencia o problemas en redacción o comprensión de los ítems previa a la aplicación masiva del instrumento (Boparai, & Kathuria, 2018). Con respecto a la validación por jueces, se considera como un componente fundamental de la evaluación de un instrumento, porque permite mediante un examen de los expertos la evaluación de los ítems de Quest (Madsø, Pachana, & Nordhus 2023). En esta fase, los jueces evalúan la claridad, relevancia y congruencia de los ítems, y verifican que el cuestionario cumpla con los requisitos de su objetivo; con respecto a este proceso, es mejor contar con especialistas en el tema y en la metodología a utilizar (Maldonado-Suárez & Santoyo-Telles 2024). Entre los métodos estadísticos que pueden aplicarse para estimar el grado de consenso entre los jueces, se encuentran el coeficiente de concordancia de Kendall y el índice de validez de contenido de Lawshe (Rodríguez et al, 2024).

En cuanto a la validez de respuesta: se asocia con la eficiencia con que el instrumento en cuestión logra captar la respuesta de los entrevistados; esta permite comprobar si los ítems son homogéneamente interpretados y si las respuestas brindadas son precisas con respecto a la información que se desea obtener (Tsang & Terkawi 2017). Además, el análisis de las preguntas abiertas podría revelar dimensiones importantes sobre las percepciones de los encuestados acerca del instrumento (Sacomori, et al, 2023).

La verificación de la respuesta de validez de un instrumento puede hacerse a través de la evaluación de la consistencia interna utilizando coeficientes como el α de Cronbach o ω , los cuales cuantifican la fiabilidad de los ítems de la escala (Polit, & Beck, 2008). Un ítem que presenta baja relación con el total del cuestionario sugiere que existen problemas en su construcción o su relación con el constructo de interés. Por esta razón, la validez interna es considerada un aspecto crítico en la determinación de los parámetros de un instrumento de medición.



Por lo anterior, para asegurar una validación rigurosa, es mejor utilizar una combinación de enfoques cualitativos y cuantitativos, involucrar a las partes interesadas relevantes y realizar pruebas preliminares destinadas a asegurar la calidad del instrumento antes de su aplicación a gran escala (Boparai, & Kathuria, 2018). Además, en estudios multiculturales, la adaptación transcultural del instrumento es esencial para mantener su validez y fiabilidad en diferentes contextos lingüísticos y culturales (Aljehani, et al, 2023).

Fase Cuantitativa: Validez Externa del Instrumento

La validez externa de un instrumento de medición se refiere a su capacidad de producir resultados generalizables, no se limita al contexto workshop en el que fue elaborado y usado. Un instrumento con alta validez externa puede ser usado en otras poblaciones, tiempos y escenarios sin que su precisión y relevancia afecten sus mediciones (Polit, & Beck, 2008). Para valorar esta característica, resulta importante investigar la estabilidad (reproducibilidad), validez de criterio (concurrente y predictiva), validez de rendimiento. Estos factores permiten asegurar que el instrumento no pierde funcionalidad y exactitud en diferentes situaciones de uso (Fenn, Tan & George 2020).

Estabilidad (Reproducibilidad): es un criterio fundamental de validez externa, pues determina la capacidad de un instrumento para obtener resultados convergentes en diferentes momentos, diferentes muestras de una población o diferentes evaluadores; un instrumento es estable si sus mediciones son homogéneas en condiciones parecidas y en un contexto que genera errores de medición por diversas condiciones contextuales (Madsø, Pachana, & Nordhus 2023).

Para determinar la estabilidad de un instrumento, se aplica confianza test-retest, en donde se aplica un test de manera temporal a la misma muestra y se observa la evolución de los resultados. Si se obtiene un coeficiente de correlación intraclase alto o un coeficiente de correlación de Pearson alto, se puede afirmar que el instrumento posee alta estabilidad y, por ende, es estable en el tiempo y en su uso (Sacomori, et al, 2023). Otra estrategia para evaluar la estabilidad es la consistencia interevaluadores en la que, diversos especialistas evalúan el mismo instrumento en diferentes evaluaciones a los mismos sujetos, busca determinar si sus evaluaciones y resultados son efectivas en gran parte coincidentes (Mondragon-Sanchez, Landeros-Olvera & Pérez-Noriega 2020).



Validez de Criterio: busca en qué cantidad un instrumento o test se correlaciona con otra medición que se considera un “test de oro” por tener validez comprobada. Dicho proceso se encarga de verificar si el instrumento que se elabora efectivamente mide el fenómeno en cuestión a la luz de otros métodos (Bautista-Espinel et al, 2022). La validez de criterio se divide en dos tipos, la primera es la concurrente y la otra predictiva.

La validez de criterio concurrente se relaciona con el uso de un instrumento en conjunto con una medida que ha sido validada en otro contexto al mismo momento. Si los dos instrumentos muestran una correlación significativa, se puede aceptar que el cuestionario o escala evaluada puede tener alto grado de validez (Ranganathan 2022). Esta validación se relaciona con el uso de adaptaciones en diferentes contextos culturales, como la validación del cuestionario CADE-Q SV en Arabia Saudita, donde se contrastó con otras herramientas de renombre para medir el conocimiento sobre la enfermedad coronaria (Aljehani, et al, 2023).

Por otro lado, la validez predictiva se centra en evaluar si el instrumento en cuestión tiene el potencial de predecir ciertos resultados en el futuro relacionados con un fenómeno particular. Se analiza si las mediciones realizadas pueden predecir de manera precisa el rendimiento o la ocurrencia de eventos que están relacionados con el constructo que se evalúa (Maldonado-Suárez & Santoyo-Telles 2024). Este tipo de evidencia de validez es particularmente importante respecto a los instrumentos diseñados para evaluar factores de riesgo, como es el caso de aquellos destinados a predecir el control del riesgo de obesidad en adolescentes mexicanos. Aquí, demostrar una fuerte validez predictiva fue crucial para la efectividad del instrumento para futuras poblaciones (Navarro-Rodríguez, et al, 2023). Para evaluar la validez predictiva se aplican coeficientes de correlación de Pearson o Spearman, análisis de regresión y, en algunos casos, modelado de ecuaciones estructurales, según el tipo de datos y nivel de medida del constructo (Tavakol, & Wetzel 2020).

Validez de Rendimiento: la validez de rendimiento se centra en el perfeccionamiento de las herramientas que permiten medir el desempeño de las personas en sus actividades laborales o en el aprendizaje de destrezas que se relacionan con el constructo en cuestión (Rodríguez et al, 2024). Es común en educación, atención médica y aquellas áreas de la psicología donde medir el progreso o la efectividad de una intervención es importante.



Esta forma de validez se establece mediante la comparación de puntuaciones derivadas del instrumento con indicadores de desempeño reales, como pruebas estandarizadas, evaluaciones de habilidades prácticas o registros clínicos (López-Belmonte, 2022). En cuanto a la confirmación de los resultados de la evaluación del rendimiento, puede haber procesos como ANOVA, regresión y pruebas de sensibilidad y especificidad, entre otros, dependiendo del tipo de datos que se haya recopilado. Un instrumento de alto rendimiento sería aquel que clasifica con precisión a los sujetos como de alto y bajo rendimiento y clasifica de manera precisa, sin sesgo y libre de errores de clasificación (Jang, Choi, & Kim 2022).

La validez externa es un aspecto importante en la creación de herramientas de medición porque define la relevancia y posible uso del cuestionario o escala en otras situaciones. La reproducibilidad o estabilidad de los resultados es una función del tiempo y las diferentes condiciones de la aplicación de sus resultados. Tanto la validez de criterio sea concurrente o predictiva, permite comprobar la medida comparándola con otras aceptadas, asegurando validez y eficacia diagnóstica y predictiva. La validez de rendimiento, por su parte, analiza si el instrumento captura adecuadamente el desempeño de los sujetos respecto al constructo en cuestión. La combinación de estos métodos asegura que el instrumento no solo sea válido en contexto de creación, sino pueda aplicar en otras investigaciones y se obtengan resultados consistentes y representativos (Polit, & Beck, 2008).

Índices Para la Evaluación de la Validez de Contenido

La validez de contenido es uno de los enfoques más comunes para verificar si un instrumento mide el dominio o la variable que se está estudiando. Para examinarlo, se utilizan tanto métodos cuantitativos como cualitativos, entre los que se encuentran el *Índice de Validez de Contenido* y el *coeficiente V de Aiken*, que determinan la relevancia, representatividad y precisión de los ítems a través de las opiniones de evaluadores expertos (Maldonado-Suárez & Santoyo-Telles 2024)

Herramientas de Medición: Fiabilidad y Consistencia Interna

En cualquier validación de instrumentos de medición, la fiabilidad y consistencia interna son aspectos críticos para la veracidad y la estabilidad de las mediciones obtenidas. Como parte de estos criterios, la fiabilidad hace referencia al seguimiento que se tiene en diferentes aplicaciones de un instrumento, de sus resultados en repetidas aplicaciones, y la variación obtenida.



Por su parte, la consistencia interna comprueba que, dentro de un cuestionario, todos los ítems que se han incluido cumplen el mismo propósito, al menos a un grado aceptable (Polit, & Beck, 2008). En el caso de los cuestionarios se estima la fiabilidad a partir de cálculos polinómicos que se pueden aplicar a los datos, o se asocia la puntuación obtenida mediante el coeficiente alfa de Cronbach, Kuder-Richardson (KR-20) o coeficiente omega de McDonald, todos estos derivados de las escalas mencionadas, para su peculiaridad en su uso (Tavakol, & Wetzel 2020).

Índices de Fiabilidad

Alfa de Cronbach: es uno de los métodos más utilizados para analizar la confiabilidad de los instrumentos, sobre todo cuando se trata de escalas tipo Likert o ítems que miden un mismo constructo en diferentes dimensiones (Boparai, & Kathuria, 2018). Este coeficiente se obtiene sobre la base de la correlación promedio entre los ítems y proporciona una medición de la homogeneidad del instrumento. Su valor se encuentra entre 0 y 1, un alfa de Cronbach en escalas superiores de 0.70 es considerado aceptable y un 0.80 indica alta confiabilidad (Ranganathan 2022). El alfa de Cronbach se ha utilizado ampliamente en estudios de validación de instrumentos en diversas disciplinas, como en la adaptación y validación del cuestionario CADE-Q SV para evaluar el conocimiento sobre la enfermedad coronaria en Arabia Saudita (Aljehani, et al, 2023). Sin embargo, este coeficiente tiene limitaciones, ya que asume que todos los ítems tienen una contribución similar a la escala y que las correlaciones entre los ítems son uniformes (Tavakol, & Wetzel 2020).

Kuder-Richardson (KR-20): se trata de una variante del alfa de Cronbach diseñada específicamente para evaluar la fiabilidad de pruebas o cuestionarios con ítems dicotómicos, en los que las respuestas se expresan en dos categorías, por ejemplo, "correcto/incorrecto" o "sí/no" (Polit, & Beck, 2008). Su aplicación es común en instrumentos de evaluación educativa y en pruebas psicométricas en las que las respuestas se categorizan en valores binarios (Fenn, Tan & George 2020). Su uso ha sido reportado en estudios de validación de instrumentos educativos y clínicos, como en la validación del EMCC-14 para evaluar la competencia cultural en profesionales de salud en Colombia (Bautista-Espinel et al, 2022).

Omega de McDonald: es una alternativa más robusta al alfa de Cronbach, especialmente en escalas con estructuras factoriales complejas o cuando los ítems no cumplen con el supuesto de tau-equivalencia (Tavakol, & Wetzel 2020).



A diferencia del alfa de Cronbach, que supone una varianza homogénea en los ítems, el coeficiente omega permite mayor flexibilidad al considerar la carga factorial de cada ítem.

Este coeficiente se considera más apropiado en estudios donde el análisis factorial confirma que los ítems se agrupan en una estructura multifactorial o cuando existen correlaciones desiguales entre los ítems (Fenn, Tan & George 2020). Investigaciones recientes han aplicado estas metodologías a la apreciación de la fiabilidad en proyectos de adaptación transcultural de herramientas de evaluación en salud y educación (López-Belmonte, et al, 2022); de hecho la valoración de la confiabilidad, y en especial de la consistencia interna, constituye un componente básico y decisivo en el proceso de validación de un instrumento, puesto que tales evaluaciones aseguran que la información recolectada sea exacta y susceptible de replicación en diferentes contextos.

Estrategias para Mejorar la Validez de Constructo

La validez de constructo es uno de los criterios más relevantes de un instrumento de medición, porque se determina en qué medida los ítems reflejan el nivel apropiado de la teoría a conceptos evaluativos (Polit, & Beck 2008). Para lograr este tipo de validez, se deben realizar varios pasos entre los más relevantes se encuentran la medición de la variabilidad y correlación en los ítems, la inversión de reactivos negativos, y el diseño de matrices de datos que optimen la información. La variabilidad y correlación de ítems es una de las técnicas más poderosas para asegurar que cada uno de los ítems del instrumento pueda discriminar efectivamente entre diferentes niveles de la variable que se está midiendo. El cálculo de la correlación y el alfa de Cronbach ayudan a identificar ítems con baja correlación con la escala total (Boparai, & Kathuria, 2018). Un ítem que tiene una baja correlación con la puntuación total es poco probable que esté discriminando adecuadamente entre los participantes o que esté midiendo el constructo apropiadamente. En tales casos, es aconsejable ajustar la formulación del ítem o eliminarlo para mejorar la consistencia interna del instrumento (Tavakol, & Wetzel 2020). Otro procedimiento necesario es el análisis factorial exploratorio y confirmatorio (AFE y AFC), examinan la integración de ítems y su relación con las dimensiones teóricas del constructo (Ranganathan 2022). Estos análisis ayudan a establecer si los encuestados agrupan los ítems en factores que representan adecuadamente el fenómeno estudiado, mejorando así la validez de constructo del instrumento (Lira & Caballero 2020).



El diseño de los instrumentos de medición recomienda la inclusión de ítems formulados en dirección opuesta, o desfavorable. Esta táctica ayuda a disminuir el sesgo en las respuestas y fortalece la validez de constructo que se está evaluando al contrarrestar el efecto de la acquiescencia, que es la tendencia generalizada de los participantes del estudio a responder a preguntas sin una consideración reflexiva (Maldonado-Suárez & Santoyo-Telles 2024).

Para llevar a cabo la inversión de ítems, se cambian las formulaciones originales de frases de afirmativas a negativas o viceversa, de modo que las respuestas se alineen más exactamente con la evaluación del constructo (Jang, Choi, & Kim 2022). Durante la fase de análisis de datos, hay un paso para recategorizar los ítems invertidos para asegurarse de que sus valores sean correctamente interpretados dentro de la escala general (Mondragon-Sanchez, Landeros-Olvera, & Pérez-Noriega 2020). Después de un proceso de inversión de ítems, se puede construir la matriz de datos. Esta matriz debe estar dispuesta de manera que cada fila sea un participante, y cada columna sea un ítem del instrumento. Esta disposición es útil para la aplicación de algunas técnicas estadísticas avanzadas, como el análisis factorial, el cálculo de la consistencia interna y el modelado de ecuaciones estructurales para probar la validez del instrumento (Rodríguez et al, 2024).

Validez de constructo (Análisis Factorial Exploratorio y Confirmatorio)

El análisis factorial es una técnica estadística utilizada en la validación de instrumentos para analizar datos con el fin de detectar subestructuras, así como verificar si los ítems se agrupan en factores que son congruentes con el constructo teórico (Tavakol, & Wetzel 2020). Se divide en análisis factorial exploratorio (AFE) (Tabla 1) y análisis factorial confirmatorio (AFC) (Tabla 2). El primero es un método que permite el descubrimiento de patrones latentes en los datos sin imponer límites a priori, mientras que el segundo se utiliza para evaluar si una estructura factorial hipotetizada es confirmada por los datos recopilados, de ahí que la elección de métodos para la extracción de datos y rotación, así como la evaluación de la adecuación de los datos, son esenciales para determinar la validez de los factores identificados; estas elecciones están influenciadas por la escala de medición, el tamaño de la muestra y la correlación entre los ítems (Ranganathan 2020; Polit, & Beck, 2008).



Tabla 1. Valores esperados en el modelo de AFE

Criterio	Valor recomendado	Interpretación
Kaiser-Meyer-Olkin (KMO)	> 0.70	Mide la adecuación de la muestra para el análisis factorial. Valores mayores indican que el análisis es adecuado.
Prueba de Esfericidad de Bartlett	$p < 0.05$	Indica que las correlaciones entre variables son lo suficientemente grandes para aplicar un análisis factorial.
Comunidades Iniciales	> 0.30	Evalúa qué porcentaje de la varianza de cada variable está explicado por los factores. Valores bajos indican variables poco representativas.
Autovalores (Eigenvalues)	> 1	Utilizado en el criterio de Kaiser, indica qué factores tienen una varianza explicativa significativa.
Varianza Total Explicada	> 50%-60%	Representa la proporción de la varianza total explicada por los factores retenidos.
Carga Factorial	> 0.30 (mínima), > 0.50 (buena), > 0.70 (muy fuerte)	Indica la fuerza de la relación entre una variable y un factor. Valores más altos reflejan mejor relación.
Rotación Factorial	Factorías deben ser interpretables	Asegura que los factores retenidos sean más fáciles de interpretar. Métodos como Varimax u Oblimin se eligen según las correlaciones entre factores.

Fuente: elaboración propia.

Tabla 2. Valores esperados en el modelo de AFC

Criterio	Valor recomendado	Interpretación
Chi-cuadrada (χ^2)	$p > 0.05$	Un valor no significativo indica que el modelo ajusta bien a los datos observados.
Razón χ^2 /Grados de Libertad	< 3 (bueno) o hasta 5 (aceptable)	Una razón baja indica un buen ajuste relativo del modelo considerando el tamaño de muestra.
Índice de Bondad de Ajuste (GFI)	> 0.90	Indica qué tan bien el modelo reproduce la matriz de covarianzas observada.
Error Cuadrático Medio de Aproximación (RMSEA)	< 0.05 (bueno), 0.05-0.08 (aceptable), > 0.10 (pobre ajuste)	Evalúa cuánto se ajusta el modelo al nivel poblacional.
Raíz Cuadrada Residual (RMR)	< 0.05	Representa la discrepancia media entre las matrices observada y estimada.
Índice Comparativo de Ajuste (CFI)	> 0.90	Compara el modelo propuesto con uno nulo, indicando qué tan bien ajusta el modelo propuesto.
Índice de Tucker-Lewis (TLI)	> 0.90	Similar al CFI, penaliza la complejidad del modelo y premia la parsimonia.

Fuente: elaboración propia.

Métodos de Extracción

En el contexto del análisis factorial, los procedimientos de extracción se utilizan para revelar los factores latentes que sustentan la covarianza observada entre variables, y su elección depende de los supuestos estadísticos formulados sobre las relaciones entre los ítems. Entre las aproximaciones más prevalentes figura el análisis de componentes principales (ACP), que se concibe como una técnica de reducción de



dimensionalidad destinada a conservar la mayor parte de la información original contenida en el conjunto de datos. En este planteamiento, los factores obtenidos denominados componentes principales son seleccionados de modo que maximicen la varianza explicada, lo que permite representar el total de datos mediante un número inferior de variables (Tavakol, & Wetzel 2020). A pesar de que el ACP no se clasifica formalmente como un análisis factorial, se aplica con frecuencia en la fase exploratoria de la investigación como un análisis preliminar que proporciona una aproximación a la estructura interna de un cuestionario o instrumento (Lira & Caballero 2020)

El Análisis Factorial Exploratorio (AFE) tiene como finalidad revelar patrones que no emergen inmediatamente en un conjunto de datos y agrupar ítems en factores que reflejan distintas dimensiones de un constructo subyacente. A diferencia del Análisis de Componentes Principales, el AFE permite postular la existencia de una estructura factorial latente y optimiza la correspondencia de los ítems a los factores comunes que a posteriori se extraen (Boparai, & Kathuria, 2018). Este procedimiento reviste una importancia crítica en la validación de cuestionarios y escalas, ya que identifica ítems redundantes, irrelevantes o que se desvinculan de forma insatisfactoria del constructo que se indaga (Ranganathan 2022).

Los Factores Principales (Principal Axis Factoring, PAF) constituyen una metodología alternativa al ACP orientada a extraer exclusivamente la varianza común entre ítems, omitiendo la varianza que se atribuye a factores específicos y a errores aleatorios. Este enfoque es particularmente útil cuando los ítems de un cuestionario se interpretan como medidas de constructos latentes que interactúan entre sí mediante otros ítems orientados al mismo fenómeno (Mondragon-Sanchez, Landeros-Olvera, & Pérez-Noriega 2020). La implementación de un PAF tiende a perfeccionar la validez de la representación de la estructura factorial en la mayoría de los contextos, convirtiéndose en un recurso particularmente ventajoso en el estudio de variables psicológicas y sociales (Aljehani, et al, 2023)

En el caso de la Estimación por Máxima Verosimilitud (MV) esta estrategia maximiza la obtención de los factores a través de la valoración de parámetros que, bajo especificación probabilística, representan de la mejor manera la estructura subyacente de los datos. Su fortaleza reside en el sólido desempeño de las pruebas asociadas y en la posibilidad de evaluar, de manera rigurosa, la adecuación del modelo factorial propuesto; por estas características, resulta especialmente pertinente en contextos de análisis



confirmatorio (Polit, & Beck, 2008). Asimismo este método asume la normalidad multivariante de los datos y es popular en la validación psicométrica de escalas (Tavakol, & Wetzel 2020).

Ahora bien, el Análisis de Componentes Principales (ACP) busca maximizar la varianza explicada y la reducción de la dimensionalidad del conjunto de variables. Este método se basa en mediciones que se realizan sin concepto de error, por lo que el Análisis Factorial Exploratorio (AFE) que considera varianza compartida, varianza específica y error es más útil ya que realiza la tarea de revelar la estructura latente de una variable sin una hipótesis previa del número de factores. Otra técnica que se enuncia es el Factores Principales (Principal Axis Factoring, AFP) que se ocupa solo de la varianza compartida y se fundamenta en la existencia de un concepto fuerte entre variables; finalmente MFVL (Método de Máxima Verosimilitud) opera bajo la premisa de que la extracción de factores y las evaluaciones del ajuste del modelo se realizan con el uso de los modelos estadísticos subyacentes; este último tiene una restricción relacionada con la distribución normal multivariante de los datos (Aljehani, et al, 2023; Fenn, Tan & George 2020; Ranganathan 2022; Tavakol, & Wetzel, 2020).

Finalmente, resulta necesario evaluar previamente la adecuación de datos mediante índice de Kaiser-Meyer-Olkin (KMO) y prueba de esfericidad de Bartlett, siendo deseable que esta última adquiera significación con un $p < .05$; confirma que la estructura de correlación satisface las exigencias para la extracción de factores. En síntesis, la selección meticulosa y aplicación de los métodos de extracción conducen a una identificación fidedigna de las estructuras latentes del conjunto de datos, lo que, a su vez, enriquece la interpretación y validez de los constructos en la literatura académica e investigaciones posteriores (Aljehani, et al, 2023; Fenn, Tan & George 2020; Ranganathan 2022; Tavakol, & Wetzel, 2020).

Métodos de Rotación

Después de extraer los factores, se aplica una técnica de rotación para mejorar la interpretación de los resultados. La rotación ayuda a optimizar la estructura factorial al mejorar la carga de los ítems en un factor particular, lo que ayuda a identificar las dimensiones más profundas (Boparai, & Kathuria, 2018).

Rotación Ortogonal: asume que los factores son independientes entre sí y no presentan correlaciones.

Este enfoque es útil en casos donde se buscan dimensiones bien definidas dentro del constructo Fenn, Tan & George 2020).



Algunos de los métodos más conocidos son: *Varimax*: aumenta la varianza compartida entre factores, lo que ayuda en la interpretación, ya que las cargas factoriales se distribuyen más claramente (Tavakol, & Wetzel 2020). *Equamax*: equilibra la distribución de las cargas factoriales entre los ítems y factores, resultando en una distribución uniforme de las cargas factoriales (Ranganathan 2022). *Quartimax*: reduce el número de factores en los que cada ítem carga significativamente, permitiendo la identificación de dimensiones generales amplias dentro del instrumento (Polit, & Beck, 2008). *Rotación Oblicua*: permite la correlación entre factores, lo que es ideal para analizar constructos psicológicos o sociales que son interdependientes por naturaleza (Mondragon-Sanchez, Landeros-Olvera, & Pérez-Noriega 2020).

Algunos de los métodos más utilizados son: *Promax*: esto es beneficioso para la interpretación al trabajar con un gran número de factores porque permite correlaciones moderadas entre los factores (Madsø, Pachana, & Nordhus 2023). *Oblimin*: proporciona una solución factorial en la que los factores pueden estar interrelacionados. Esto es útil para la investigación en la que se espera que las relaciones entre dimensiones del constructo interactúen (Ranganathan 2022).

La rotación en el análisis factorial se organiza con el propósito de hacer más evidente la estructura de las cargas factoriales, al acentuar los patrones latentes en la matriz de datos. Al variar las asignaciones de las variables a los factores, se reposicionan las conexiones para que las relaciones entre ellas surjan de forma más nítida (Polit, & Beck, 2008). La elección del método de rotación más apropiado aconseja, en primera instancia, el examen de la matriz de correlación de los factores emergentes. Correlaciones que superan 0,3 sugieren la conveniencia de una rotación oblicua, que permite reflejar la interdependencia de las dimensiones; cuando las correlaciones no alcanzan significación, se justifica el empleo de una rotación ortogonal; además, el nivel de medición y la naturaleza de los datos son cruciales, ya que afectan la eficiencia de la interpretación factorial (Lira & Caballero 2020; Maytoarena, Reyes, & Sánchez 2023).

Análisis de Reproducibilidad en Instrumentos

La reproducibilidad es la base de la confiabilidad de la cuantificación y evaluación de fenómenos en un espacio de tiempo. Para el caso de los instrumentos cuantitativos, se podría decir que se analiza mediante el test-retest que consiste en la administración del mismo instrumento a la misma muestra en



dos diferentes momentos. La reproducibilidad se mide con coeficientes de correlación que son el CCI y el pearson, donde usualmente, el CCI mayor de 0.80 se considera buena reproducibilidad del instrumento (Boparai, & Kathuria, 2018; Tavakol, & Wetzel 2020).

La estabilidad también se puede analizar a partir de la consistencia inter evaluadores, donde varios evaluadores apliquen el utensilio a la misma muestra en un intento de verificar si los resultados obtenidos son los mismos en todos los casos. Este procedimiento es muy útil en las situaciones donde se llevan a cabo mediciones o evaluaciones subjetivas, como las evaluaciones clínicas o educativas (Sacomori, et al, 2023). En los instrumentos cualitativos, la reproducibilidad se aborda mediante triangulación de datos, que consiste en obtener respuestas de distintas fuentes para buscar patrones comunes que respalden la confiabilidad de los resultados (Maldonado-Suárez & Santoyo-Telles 2024). Una técnica importante es el análisis de concordancia entre evaluadores, que compara diferentes versiones de los mismos datos cualitativos que los investigadores han interpretado de forma independiente (Fenn, Tan & George 2020).

El análisis de reproducibilidad es esencial en estudios de validación de instrumentos de distintas disciplinas. Es el caso de la validación del cuestionario EMCC-14, donde se utilizó test-retest para medir la estabilidad de las mediciones en un grupo de profesionales de la salud (Bautista-Espinel et al, 2022). Igualmente, en investigaciones de corte cualitativo sobre conciencia en salud, la triangulación de datos ha sido fundamental en diferentes poblaciones para confirmar la estabilidad de los hallazgos (Apodaca-Armenta, Gómez-Rodríguez, & Rodríguez-Vázquez 2024).

Validez de Criterio (Concurrente y Predictiva)

Validez Concurrente: relación con medidas preexistentes. La validez concurrente estudia la relación entre el instrumento en evaluación con otro validado que mide el mismo constructo en su forma sincronizada. Este procedimiento da la posibilidad de determinar si el nuevo instrumento genera resultados que son congruentes con los obtenidos a través de instrumentos preexistentes en su nivel de fiabilidad, lo que posibilita la normatización en el ámbito de su uso (Madsø, Pachana, & Nordhus 2023). Para determinar la validez concurrente, se utilizan correlaciones como el coeficiente Pearson en escalas de intervalo o razón, y el coeficiente de Spearman en escalas ordinales (Tavakol, & Wetzel 2020). Un valor de correlación mayor a 0.70 establece que existe relación importante entre los dos instrumentos,



lo que sugiere que la herramienta evaluada es apropiada para el constructo en cuestión (Polit, & Beck, 2008).

Un ejemplo reciente de validación concurrente es la adaptación y validación del cuestionario CADE-Q SV, en el cual se evaluaron sus resultados en relación a otro cuestionario validado sobre el conocimiento de enfermedades cardiovasculares en Arabia Saudita (Aljehani, et al, 2023). También, en la validación del cuestionario Quality of Recovery-15, se empleó un método análogo que buscó correlacionar el cuestionario con estandarizadas la postoperatoria de la recuperación clinical scales (Echeverri-Mallarino, & Rodríguez Romero 2023).

validez predictiva: la validez predictiva se centra en la relación entre un instrumento y la acción o comportamiento de los individuos en relación con el constructo dado el marco temporal de la evaluación. A diferencia de la validez concurrente, estas formas de validez requieren la implementación del instrumento en algún momento y luego comparar los resultados con mediciones posteriores que representan el resultado de interés (Boparai, & Kathuria, 2018).

Las metodologías más frecuentemente empleadas para la evaluación de la validez predictiva comprenden, en primer término, el análisis de regresión lineal y logística, cuyo objeto es determinar hasta qué punto los puntajes obtenidos mediante el instrumento pronostican resultados futuros en variables dependientes determinadas (Tavakol, & Wetzel 2020). Complementariamente, se recurre al coeficiente de determinación (R^2), que refleja el porcentaje de la variabilidad observada en la variable criterio que puede ser atribuida al instrumento que se está considerando (Polit, & Beck, 2008). Asimismo, la curva ROC (Receiver Operating Characteristic) proporciona una medida de la capacidad para identificar, con precisión, eventos clínicamente relevantes a futuro y, mediante el cálculo de sensibilidad y especificidad, valora el rendimiento del instrumento (Sacomori, et al, 2023). Un caso ilustrativo de validez predictiva es el trabajo de Navarro-Rodríguez et al., en el que los investigadores se propusieron determinar la capacidad de una escala para anticipar el riesgo de obesidad en adolescentes mexicanos; el análisis longitudinal mostró que las puntuaciones obtenidas en el cuestionario se asociaron de forma robusta a la incidencia de obesidad durante el periodo de seguimiento, confirmando, en consecuencia, la utilidad del instrumento como herramienta de detección en el contexto poblacional considerado.



La validez de criterio en sus formas concurrente y predictiva es esencial en el proceso de validación de cualquier herramienta de medición; la primera se valida en correlación con otras herramientas de medición que ya están validadas, y la última se enfoca en qué tan bien la herramienta predice eventos o comportamientos futuros. Ambos métodos de validez involucran la aplicación de rigurosos métodos estadísticos como la correlación, regresión y análisis de la curva ROC, dependiendo del contexto específico. Estudios han evidenciado cómo estos métodos son aplicables en contextos de salud y psicología, así como en educación e inteligencia artificial (Ranganathan 2022); concluyendo que el análisis exhaustivo de estos factores ayuda a mejorar el enfoque y la amplitud de la herramienta de medición. Esto asegura la validez de la herramienta en múltiples contextos de investigación.

Evaluación de la Efectividad del Instrumento. Medir la efectividad de un instrumento requiere un proceso multifásico que determine la consistencia y la capacidad de éste para capturar la dimensión que le corresponde. Para ello, resulta imprescindible la práctica sistemática de tests que indaguen tanto la fiabilidad interna como la validez y la adecuación de la herramienta. La inclusión de procedimientos estadísticos, destacando el análisis factorial, se convierte en un paso operativo que revela la jerarquía estructural y el grado en que cada ítem se vincula coherentemente con la construcción general (Boparai, & Kathuria, 2018; López-Belmonte, 2022). Complementariamente, la fiabilidad puede fortalecerse mediante la realización de estudios piloto y el cálculo de correlaciones entre ítems, procedimientos que permiten ajustar la red de indicadores y reducir la probabilidad de sesgos en los datos (Polit, & Beck, 2008). El conjunto de estas prácticas, tal como evidencian investigaciones recientes, se manifiesta como una garantía de robustez metodológica que sustenta tanto los proyectos de investigación como los procedimientos de evaluación en el ámbito profesional.

Definición del Punto de Corte Óptimamente. Un límite de decisión óptimo evita al máximo los conflictos que deben ser clasificados por el sistema. La sensibilidad y especificidad del instrumento en la evaluación permiten establecer un nivel que minimiza la proporción de falsos positivos y falsos negativos. El índice de Youden se utiliza a menudo para encontrar el punto de equilibrio entre sensibilidad y especificidad (Ranganathan 2022). La precisión en el diagnóstico y la minimización de errores en la clasificación sin duda permiten mejorar el uso y la utilidad en el contexto clínico y en el contexto de investigación.



La curva ROC (Receiver Operating Characteristic) es una de las herramientas más fundamentales para evaluar la efectividad de un instrumento, ya que ilustra la relación entre la sensibilidad y la tasa de falsos positivos (1 - especificidad). El gráfico ilustra el rendimiento general del instrumento en varios umbrales y su precisión al diagnosticar la condición. El rendimiento se captura utilizando el área bajo la curva (AUC), que es una medida de efectividad. Los valores más altos de AUC (más cercanos a 1) indican una mejor capacidad de discriminación entre las diferentes condiciones o estados a evaluar (Maytorena, Reyes, & Sánchez 2023).

CONCLUSIONES Y RECOMENDACIONES

Para el diseño de un instrumento de medida se requiere una revisión exhaustiva, que constituye una base teórica de la cual se extraen conceptos relevantes para la investigación. En esta fase, el contenido teórico se selecciona para el diseño de cuestionarios cuyas preguntas se alineen a los objetivos de la investigación; luego, los cuestionarios son revisados por un grupo de expertos para asegurar, mediante su criterio, que se reúnan los parámetros de pertinencia, claridad y representatividad en el contenido.

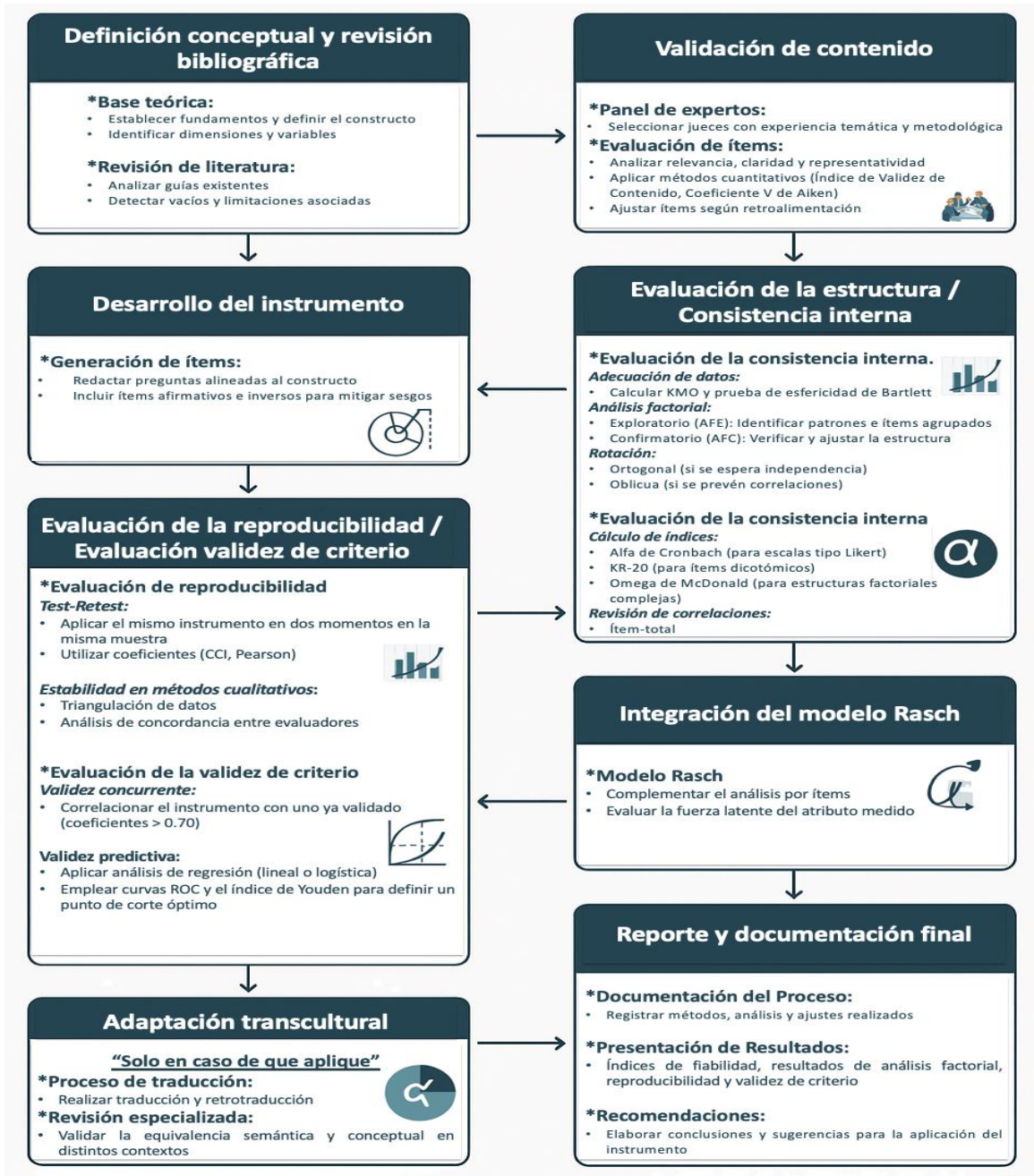
Cuando hablamos de validez de constructo. Esto se realiza mediante análisis por medio de análisis factoriales exploratorios que permiten comprender cómo se distribuyen los ítems en función de las dimensiones teóricas. Esto también permite mejorar la medición del constructo mediante la corrección de los errores y las redundancias que haya en la mezcla del constructo y se complementa con las consistencias internas de las mediciones que se denominan, el calcular valores como el alfa de Cronbach de los ítems. Posteriormente se procede a realizar la validez de criterio, herramienta capaz de demostrar la mayor capacidad y usabilidad asertiva de un instrumento de medida en la que se busca identificar mediante la utilización de un criterio externo (algo que ya es aceptado como válido o relevante). Si existe relación fuerte entre ellos, se puede inducir que el instrumento tiene buena validez de criterio. Finalmente, para aumentar la aplicabilidad contextual y poblacional diversa del instrumento, se añade un procedimiento de adaptación transcultural. Esta fase del proceso consiste en la traducción del contenido y la retrotraducción, así como en la revisión por parte de expertos que aseguran que la intención y el significado de los ítems logren coherencia en múltiples culturas e idiomas.

Para resumir, el diseño y validación de un instrumento de investigación (ver figura 1) implica: desarrollar una base teórica sólida, validación experta del contenido, realizar un análisis factorial para



confirmar la estructura interna, establecer puntos óptimos de corte altamente discriminatorios y realizar una adaptación transcultural para la relevancia en múltiples contextos. La meticulosa ejecución de estos pasos asegura que la herramienta de investigación deseada sea precisa y confiable, capaz de generar datos consistentes y comparables para la investigación.

Figura 1. Algoritmo del diseño y validación de un instrumento de medición



Fuente: elaboración propia

REFERENCIAS BIBLIOGRAFICAS

- Aljehani, R., Aljehani, G., Alharazi, H., & de Melo Ghisi, G. L. (2023). Translation, cultural adaptation and psychometric validation of the Arabic short version of the coronary artery disease education questionnaire (CADE-Q SV) in Saudi Arabia. *PEC innovation*, 3, 100205. Disponible en: <http://dx.doi.org/10.1016/j.pecinn.2023.100205>
- Al-Madaney, M. M., & Fässler, M. (2023). Development and validation of a tool to assess researchers' knowledge of human subjects' rights and their attitudes toward research ethics education in Saudi Arabia. *BMC Medical Ethics*, 24(1), 94. Disponible en: <http://dx.doi.org/10.1186/s12910-023-00968-z>
- Apodaca-Armenta, R. E., Gómez-Rodríguez, G. A., & Rodríguez-Vázquez, N. (2024). Instrumentos para medir conciencia en salud en jóvenes: una revisión sistemática. *Sanus*, 9. Disponible en: <http://dx.doi.org/10.36789/revsanus.vi1.412>
- Bautista-Espinel, G. O., Velandia-Galvis, M. L., Ortega-Cadena, N., Chaves, C. A., Ureña-Molina, M. D. P., & Gonzales-Escobar, D. S. (2022). Adaptación cultural y validación del instrumento EMCC-14, de competencia cultural en profesionales de salud colombianos. *Index de Enfermería*, 31(2), 120-124. Disponible en: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1132-12962022000200015&lng=es
- Boparai, J. K., Singh, S., & Kathuria, P. (2018). How to design and validate a questionnaire: a guide. *Current clinical pharmacology*, 13(4), 210-215. Disponible en: <http://dx.doi.org/10.2174/1574884713666180807151328>
- Echeverri-Mallarino, V., & Rodríguez Romero, V. A. (2023). Validation and cross-cultural adaptation of the quality of recovery-15 questionnaire in a Spanish-speaking population in Colombia. *BJA Open*, 8, 100231. <https://doi.org/10.1016/j.bjao.2023.100231>
- Fenn, J., Tan, C. S., & George, S. (2020). Development, validation and translation of psychological tests. *BJPsych advances*, 26(5), 306-315. Disponible en: <http://dx.doi.org/10.1192/bja.2020.33>
- González-Ramírez, P., Matus-Miranda, R., & Fernández-García, V. (2017). Factores determinantes en la formación de pasantes en servicio social. Construcción y validación de un instrumento. *Enfermería universitaria*, 14(1), 3-9. Disponible en: <http://dx.doi.org/10.1016/j.reu.2016.10.003>



- González-Ramírez, T., & García-Hernández, A. (2022). Design and Validation of a Questionnaire to Assess Student Satisfaction with Mathematics Study Materials. *International Journal of Instruction*, 15(1), 1-20. Disponible en: <https://idus.us.es/items/3bc48b91-e6f9-466f-9d54-c9e72144e338>
- Jang, Y., Choi, S., & Kim, H. (2022). Development and validation of an instrument to measure undergraduate students' attitudes toward the ethics of artificial intelligence (AT-EAI) and analysis of its difference by gender and experience of AI education. *Education and Information Technologies*, 27(8), 11635-11667. Disponible en: <http://dx.doi.org/10.1007/s10639-022-11086>
- Lira, M. T., & Caballero, E. (2020). Adaptación transcultural de instrumentos de evaluación en salud: Historia y reflexiones del por qué, cómo y cuándo. *Revista Médica Clínica Las Condes*, 31(1), 85-94. Disponible en: <http://dx.doi.org/10.1016/j.rmcl.2019.08.003>
- López-Belmonte, J., Pozo-Sánchez, S., Lampropoulos, G., & Moreno-Guerrero, A. J. (2022). Design and validation of a questionnaire for the evaluation of educational experiences in the metaverse in Spanish students (METAEDU). *Heliyon*, 8(11). Disponible en: <http://dx.doi.org/10.1016/j.heliyon.2022.e11364>
- Maldonado-Suárez, N., & Santoyo-Telles, F. (2024). Validez de contenido por juicio de expertos: Integración cuantitativa y cualitativa en la construcción de instrumentos de medición. *REIRE Revista d'Innovació i Recerca en Educació*, 17(2), 1-19. Disponible en: <https://revistes.ub.edu/index.php/REIRE/article/download/46238/41929/131290>
- Madsø, K. G., Pachana, N. A., & Nordhus, I. H. (2023). Development of the observable well-being in living with dementia-scale. *American Journal of Alzheimer's Disease & Other Dementias*®, 38, 15333175231171990. Disponible en: <http://dx.doi.org/10.1177/15333175231171990>
- Maytorena, R. S., Reyes, R. J. T., & Sánchez, S. C. R. (2023). Relevancia del diseño y validación de instrumentos de medición documental en la investigación de enfermería. *Parainfo Digital*, e35064d-e35064d. Disponible en: <https://ciberindex.com/index.php/pd/article/view/e35064d>
- Mondragon-Sanchez, E. J., Landeros-Olvera, E., & Pérez-Noriega, E. (2020). Validación de la Escala de Miedo a la Muerte de Collett-Lester en estudiantes universitarios de enfermería de México. *MedUNAB*, 23(1), 11-22. Disponible en: <http://dx.doi.org/10.29375/01237047.3723>



- Navarro-Rodríguez, D. C., Guevara-Valtier, M. C., Cárdenas-Villareal, V. M., Paz-Morales, M. D. L. Á., Urchaga-Litago, J. D., & Guevara-Ingelmo, R. M. (2023). Psychometric analysis of the obesity risk control scale in mexican young people. *Sanus*, 8. Disponible en: <http://dx.doi.org/10.36789/revsanus.vi1.337>
- Polit, D. F., & Beck, C. T. (2008). *Nursing research: Generating and assessing evidence for nursing practice*. Lippincott Williams & Wilkins. Disponible en: https://scholar.google.com/scholar?hl=es&as_sdt=0%2C5&q=Polit+DF%2C+Tatano+BC.+Measurement+and+data+quality.+En%3A+Nursing+research%3A+Generating+and+assessing+evidence+for+nursing+practice.+Wolters+Kluwer%3B+2021.+p.+320%E2%80%93931.+++&btnG=
- Ranganathan P, Designing and validating a research questionnaire - Part 1. (2022). *Perspect Clin* 14(3):152–5. Disponible en: http://dx.doi.org/10.4103/picr.picr_140_23
- Rodríguez-Rojas, G., Torres-Reyes, A., Hidalgo-Arce, I., Salazar-Mendoza, F., Pérez-Vázquez, G., & Adrián Morales-Castillo, F. (2024). Diseño, validación y confiabilidad del instrumento para prevenir infección del sitio quirúrgico. *Journal Health NPEPS*, 9(1). Disponible en: <http://dx.doi.org/10.30681/2526101012243>
- Romero, K. P., & Mora, O. M. (2020). Análisis factorial exploratorio mediante el uso de las medidas de adecuación muestral kmo y esfericidad de bartlett para determinar factores principales. *Journal of science and research*, 5(CININGEC), 903-924. Disponible en: <https://revistas.utb.edu.ec/index.php/sr/article/view/1046>
- Sacomori, C., Lorca, L. A., Martínez-Mardones, M., Pizarro-Hinojosa, M. N., Rebolledo-Díaz, G. S., & Vivallos-González, J. A. (2023). Spanish version of the ICIQ-Bowel questionnaire among colorectal cancer patients: Construct and criterion validity: Comprehensive assessment of bowel function. *BMC gastroenterology*, 23(1), 352. Disponible en: <http://dx.doi.org/10.1016/j.bjao.2023.100231>
- Supo, J. (2013). Cómo validar un instrumento. Disponible en: https://www.cua.uam.mx/pdfs/coplavi/s_p/doc_ng/validacion-de-instrumentos-de-medicion.pdf
- Tavakol, M., & Wetzel, A. (2020). Factor Analysis: a means for theory and instrument development in support of construct validity. *International journal of medical education*, 11, 245. Disponible en:



<http://dx.doi.org/10.5116/ijme.5f96.0f4a>

Torres-Hernández, E. F. (2022). Diseño y validación de un cuestionario de vocación docente para profesores (CVOC-D). *Revista Fuentes*, 24(1), 104-115. Disponible en:

<http://dx.doi.org/10.12795/revistafuentes.2022.15751>

Tsang, S., Royse, C. F., & Terkawi, A. S. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi journal of anaesthesia*, 11(Suppl 1), S80-S89. Disponible en: http://dx.doi.org/10.4103/sja.SJA_203_17

Vallejo-Gómez, L., Ruiz-Recéndiz, M., Jiménez-Arroyo, V., Alcántar-Zavala, M. L. A., & Huerta-Baltazar, M. I. (2021). Instrumentos para medir competências de enfermagem: revisão sistemática. *Sanus*, 6. Disponible en: <http://dx.doi.org/10.36789/revsanus.vi1.198>

