

Recepción: 25.10.2025

Revisión: 06.11.2025

Publicación: 08.05.2026

<https://orcid.org/0000-0002-6331-5953>

<https://orcid.org/0000-0002-7115-968X>

<https://orcid.org/0009-0008-8499-209X>

<https://orcid.org/0000-0003-4922-848X>

<https://orcid.org/0000-0001-9839-982X>

VALIDEZ Y FIABILIDAD: CLAVES PARA EL DESARROLLO DE UN CUESTIONARIO COMO INSTRUMENTO DE MEDICIÓN CONFIABLE

VALIDITY AND RELIABILITY: KEYS TO THE DEVELOPMENT A QUESTIONNAIRE AS A RELIABLE MEASUREMENT INSTRUMENT

Jazmín Zaragoza-Alonso ⁽¹⁾,
José Muñoz Flores ⁽¹⁾,
Fátima Andrea Bonilla Ramos ⁽²⁾,
María Elena Ramos Cassellis ⁽³⁾,
Jesús Andrés Arzola Flores ⁽³⁾*

⁽¹⁾ Posgrado en Ciencias Ambientales, Instituto de Ciencias, Benemérita Universidad Autónoma de Puebla, Ciudad Universitaria, Puebla, C.P. 72570.

⁽²⁾ Maestría en Manejo Sostenible de Agroecosistemas, CENAGRO, Instituto de Ciencias, Benemérita Universidad Autónoma de Puebla. Carretera a San Baltazar Tetela km 1.7, Ecocampus Valsequillo, San Pedro Zacachimalpa, Puebla, México. C.P.72960.

⁽³⁾ Facultad de Ingeniería Química, Benemérita Universidad Autónoma de Puebla. Avenida San Claudio y 18 Sur, Ciudad Universitaria, Puebla, C.P. 72570

*Autor de correspondencia:

jesus.arzolaflares@correo.buap.mx. Tel. (221) 203 8067
za224570202@alm.buap.mx, mf224570198@alm.buap.mx, br224470530@alm.buap.mx, cassellis71@yahoo.com.mx, jesus.arzolaflares@correo.buap.mx*

Resumen

La fiabilidad y la validez son elementos fundamentales en la creación de instrumentos de medición, ya que garantizan la calidad y precisión de los datos obtenidos. La fiabilidad se refiere al grado en que se obtienen resultados consistentes y estables a lo largo del tiempo. Mientras que la validez indica qué tan bien se mide lo que se pretende medir. Ambas aseguran que los resultados sean representativos y útiles para el propósito del estudio. Además, están estrechamente relacionadas: un instrumento puede ser confiable pero no válido; sin embargo, no puede ser válido si no es confiable. Por esto, es crucial evaluar diferentes tipos de validez y de fiabilidad. Este trabajo aborda la importancia de evaluar la fiabilidad y validez de un cuestionario, ya que es fundamental para obtener información precisa, comparable y de calidad. Asimismo, se presentan recomendaciones para su diseño, desarrollo y evaluación, así como descripciones de softwares y herramientas para realizar las pruebas estadísticas que permiten medir la validez y la fiabilidad.

Palabras clave: *fiabilidad, validez, cuestionario, escala psicométrica, software*

Abstract

Reliability and validity are fundamental elements in the creation of measurement instruments, as they guarantee the quality and accuracy of the data obtained. Reliability refers to the degree to which consistent and stable results are produced over time. While validity indicates how well it measures what it is intended to measure. Both ensure that the results are representative and useful for the purpose of the study. Moreover, they are closely related: an instrument can be reliable but not valid, however, it cannot be valid if it is not reliable. Therefore, it is crucial to evaluate different types of validity and reliability. This paper addresses the importance of assessing the reliability and validity of a questionnaire, as it is essential to obtain accurate, comparable, and quality information. It also provides recommendations for designing, developing and evaluating, as well as descriptions of software and tools to perform statistical tests to measure validity and reliability.

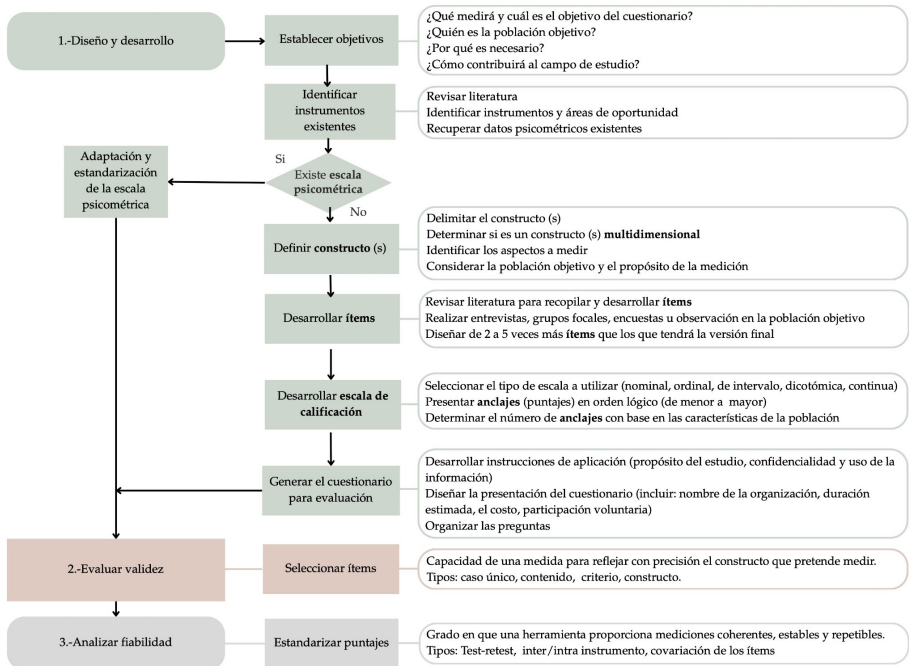
Keywords: *reliability, validity, questionnaire, psychometric scale, software*

Introducción

En cualquier disciplina que involucre la obtención de información —ya sea en ciencias sociales, salud, ingeniería o educación— es fundamental asegurarse de que el instrumento de medición realmente mida lo que debe medir y si la información recolectada es confiable y repetible (Bolarinwa, 2015). Si un cuestionario, un sensor o una escala psicométrica no mide con precisión lo que debe medir (**validez**) o si sus resultados varían sin razón aparente (**fiabilidad**), los datos recopilados pueden ser erróneos o engañosos (Phillips et al., 2021; Taherdoost, 2016). El cuestionario se ha convertido en el instrumento de recolección de datos más empleado para obtener información relevante, precisa y

confiable en distintas áreas como salud, economía, psicología, educación e investigación conductual (Bolarinwa, 2015; Taherdoost, 2016; Yusoff et al., 2021). Debido a sus múltiples ventajas permite recopilar información de una gran población con el anonimato de los participantes, con costo y tiempo menores, con preguntas estandarizadas y sin el sesgo del encuestador (Elangovan & Sundaravel, 2021). Sin embargo, si el cuestionario tiene un desarrollo deficiente, las preguntas son incompletas, imprecisas o mal redactadas, no se podrá obtener información de calidad y por ende la investigación será inadecuada (Elangovan & Sundaravel, 2021; Yusoff et al., 2021).

Figura 1 Proceso para diseñar, desarrollar y evaluar un cuestionario



Nota. Elaboración propia con base en: (Phillips et al., 2021; Swan et al., 2023; Taherdoost, 2022; Yusoff et al., 2021).

¿Cuándo se debe probar la **validez y fiabilidad** de un cuestionario? Siempre que se desarrolle uno nuevo o se adapte alguno (Bolarinwa, 2015), ya que los existentes no se ajustan a la población objetivo, no responden al problema de investigación o las propiedades psicométricas (**validez y fiabilidad**) de la escala no son consistentes (Swan et al., 2023). En muchas áreas la información recolectada sirve para tomar decisiones y ahí radica la importancia de un instrumento de medición confiable.

En ocasiones se desconoce el proceso para su diseño, así como las pruebas estadísticas que

deben ser usadas para asegurar su **validez y fiabilidad**, el software a emplear y si este es gratuito o de licencia. Resultando fundamental asegurar el correcto diseño, desarrollo y evaluación para obtener un cuestionario y datos de calidad a través de una planificación sistemática rigurosa (Elangovan & Sundaravel, 2021; Yusoff et al., 2021). Este artículo expone el proceso que debe seguirse a partir del diseño hasta su evaluación (ver figura 1.), expone qué son la validez y la fiabilidad, además de comparar distintos softwares y herramientas para su análisis.

Diseño y desarrollo de un cuestionario

Un cuestionario es un instrumento con un conjunto de preguntas, declaraciones o palabras estímulo predeterminados que permiten medir un **constructo** (una idea abstracta precisa, delimitada que es intangible pero puede medirse y analizarse) o varios a través de una **escala psicométrica** (una medida cuantitativa que mide a los participantes del estudio con relación al **constructo**) como la escala de Neofobia Alimentaria o la de Rechazo a las Alternativas Cárnicas, siempre debe tener una intención relacionada con el objetivo de la investigación y debe permitir recolectar información de calidad (Bolarinwa, 2015; Elangovan & Sundaravel, 2021; Robinson, 2018).

En la etapa de diseño y desarrollo, primero se deben establecer los objetivos del cuestionario, saber qué medirá (conceptos, características o rasgos de alguien o algo), conocer quiénes son la población objetivo (productores, estudiantes, consumidores, pacientes), determinar cuál es la necesidad de su desarrollo y cómo contribuirá en el área de estudio para evitar duplicar un instrumento existente (Yusoff et al., 2021). Si existe la **escala psicométrica**, suele ser necesario evaluar sus **propiedades psicométricas** (propiedades de medición: **validez y fiabilidad**) o adaptarla para una población o contexto cultural distinto al que se ha empleado (Swan et al., 2023).

Figura 2 Componentes de una escala psicométrica

Ítem					
Preguntas o enunciados	Escala de calificación			Puntos de respuesta	
	Totalmente en desacuerdo	En desacuerdo	Neutral	De acuerdo	Totalmente de acuerdo
1.-Me preocupa el medio ambiente	☐	☐	☐	☐	☐
2.-Proteger el medio ambiente aumenta mi calidad de vida	☐	☐	☐	☐	☐

Nota. Adaptado de Robinson (2018).

La **escala psicométrica** está conformada por ciertos elementos que tienen una función específica (Figura 2.). Comprende múltiples **ítems** (preguntas o enunciados que representan un aspecto del **constructo** que se está evaluando y que se mide cada uno con una **escala de calificación**) que generan datos paramétricos que deben cumplir con **propiedades psicométricas** (Robinson, 2018; Swan et al., 2023).

Para generar los **ítems** se recomienda hacer una revisión de literatura y considerar a actores clave de la población de estudio, esto permite una base teórica sólida, así como la inclusión de la perspectiva de las personas a investigar. Es importante considerar que la calidad de la **escala psicométrica** depende de la calidad de los **ítems** (Swan et al., 2023), en la tabla 1 se presentan recomendaciones para su diseño. Respecto al número de **ítems** Yusoff

et al. (2021) recomiendan 10 por atributo para la etapa de desarrollo, Robinson (2018) 4 y Swan et al. (2023) entre 2 y 5 veces más que la versión final.

Es importante considerar desarrollar un cuestionario que implique menos de 15 minutos para contestarse (Yusoff et al., 2021), ya que las preguntas, declaraciones y respuestas más fáciles de contestar generan los datos más concisos (Krosnick, 2018). Debido a que las preguntas son la forma en que se obtendrá la información y la calidad de las respuestas dependen de su diseño, esté debe realizarse correctamente, “las buenas preguntas son **fiabiles** (proporcionan medidas consistentes en situaciones comparables) y **válidas** (las respuestas corresponden a lo que se pretende medir)” (Fowler, 2009, p.87 en Bourke, 2016) por lo que en la Tabla 2 se presentan consejos para su redacción.

Tabla 1 Recomendaciones para redactar ítems

Recomendaciones	
✓	Redacción clara y lo más corta posible
✓	Simple y directo para vincularse con la escala de calificación
✓	Lenguaje no ambiguo, considerando el nivel educativo de la población objetivo
✓	Un concepto por ítem
✓	Redundancia o similitud de los ítems para asegurar una cobertura total del constructo
✓	La redacción debe estar libre de sesgos culturales
✓	Evitar dobles negaciones
✓	Evitar redacción negativa o codificación inversa (algunas veces se usa para evitar aquiescencia o el sesgo de respuesta extrema, siempre debe justificarse su inclusión).

Nota. Elaboración propia con base en (Dueber et al., 2022; Swan et al., 2023; Yusoff et al., 2021).

La elección de la **escala calificación o de respuesta** permite garantizar adecuadas **propiedades psicométricas** comparadas con escalas de acuerdo-desacuerdo por lo que su elección es vital para obtener información de calidad (DeCastellarnau, 2018). Además, muestra el diseño, formato, redacción y mide la respuesta de los participantes a cada pregunta o enunciado (Robinson, 2018; Swan et al., 2023). Para su diseño, debe considerarse que las escalas largas son ambiguas (Krosnick, 2018a), por lo que el número de anclajes o puntajes de respuesta recomendado es de 7 o 9.

Aunque la extensión depende de varios factores. Por ejemplo, de la población objetivo (menores para niveles educativos bajos como niños y mayores para personas con educación superior o

posgrado) (Swan et al., 2023). Del espacio de visualización de respuesta, en teléfonos inteligentes se recomiendan 5 anclajes (Robinson, 2018). Según la polaridad del constructo, si este es **bipolar** (es decir, va de un punto positivo a un negativo y tiene un punto neutral) la escala debe ser bipolar y debe tener 7 puntos. Por el contrario, si el constructo es **unipolar** (va de cero a un punto máximo, sin punto neutral) debe ser unipolar y tener 5 puntos, esto evita sesgar los datos al polo positivo o negativo y maximizar la discriminación sin comprometer la fiabilidad (DeCastellarnau, 2018; Krosnick, 2018a). También es importante considerar que la fiabilidad es mayor con escalas unipolares que bipolares (DeCastellarnau, 2018). Lo anterior muestra la relevancia de realizar una profunda revisión de literatura para la elección correcta de la escala de acuerdo con las características del estudio.

Tabla 2 Recomendaciones y ejemplos para redactar preguntas

Recomendación	Pregunta incorrecta	Pregunta correcta
✓ Las preguntas deben cumplir con el objetivo de la investigación		
✓ Incluir preguntas de filtro		
✓ Simple, directa, comprensible	Con base en su fecha de nacimiento, ¿cuál es su edad, actualmente?	¿Cuántos años tiene?
✓ Sin jerga (lenguaje informal entre un grupo de personas) o lenguaje técnico	¿Cuál es la huella hídrica de un refresco?	¿Cuánta agua se necesita para producir un refresco?
✓ Ser específico	¿Cómo fue su experiencia?	¿Cómo evaluaría la rapidez del servicio?
✓ Permitir todas las respuestas posibles	¿Cuál es la red social que usa con mayor frecuencia?	¿Cuáles son las redes sociales que usa con mayor frecuencia? Puede seleccionar varias respuestas
✓ Frecuencias y períodos deben especificarse claramente	¿Consume café orgánico frecuentemente?	¿Cuál es la frecuencia de consumo de café orgánico? A) Nunca B) 1 vez al año B) 1 vez al mes C) 1 vez a la semana D) Todos los días

Recomendación	Pregunta incorrecta	Pregunta correcta
✓ Si se busca un número (cantidades) solicitarlo como pregunta abierta exceptuando ingresos económicos	Elija la cantidad de hijos qué tiene	¿Cuántos hijos tiene?
✓ Las preguntas se leen suavemente en voz alta	¿Está casado (a) o soltero (a)?	¿Cuál es su estado civil?
Evitar:		
✓ Palabras ambiguas	¿Qué chatarra consumes todos los días?	¿Qué alimentos chatarra consumes todos los días?
✓ Preguntas de doble filo	¿Está satisfecho con el sabor del café y con el precio?	¿Cuál es el nivel de satisfacción con el sabor del café? ¿Cuál es el nivel de satisfacción con el precio del café?
✓ Negaciones	¿No crees que uno de los motivos de consumir café orgánico es por ser bueno para el ambiente?	¿Cuáles son los motivos por los que consumes café orgánico?
✓ Preguntas capciosas	¿Qué cantidad de café debes tomar al día para que tu cuerpo no se envenene?	¿Qué cantidad de café consumes todos los días?
✓ Palabras emocionalmente cargadas	¿Cómo te sientes al saber que millones de animales son sacrificados cruelmente para el consumo de carne?	¿Cuál es tu opinión sobre el consumo de carne?
✓ Nombres de prestigio	¿Qué tan saludable consideras el yogur Chobani?	¿Qué tan saludable consideras el yogur griego?
✓ Tener las opciones "otro" o "algo más" en las respuestas	¿Dónde compras alimentos saludables? A) Mercado B) Supermercado C) En línea D) Otro	¿Dónde compras alimentos saludables? A) Mercado B) Supermercado C) En línea

Nota. Elaboración propia con base en Krosnick (2018a, 2018b) y Taherdoost, (2022).

Una vez desarrollada la escala de calificación, es necesario generar la estructura y las secciones del cuestionario para proceder a la evaluación de las propiedades psicométricas. Este suele estructurarse en tres secciones: una administrativa, en la que se incluyen las instrucciones de aplicación, información de distribución del cuestionario o entorno de aplicación y no suele incluirse en el

cuestionario, la segunda tiene que ver con el orden de las preguntas y la última con datos sociodemográficos (Dalati & Marx Gómez, 2018). Es relevante considerar el orden de las preguntas, apesar de que este no afecta el significado de los cuestionamientos, si el proceso cognitivo de quienes responden, por lo que en la tabla 3 se presentan recomendaciones para su ordenamiento (Taherdoost, 2022).

Tabla 1 Recomendaciones para redactar ítems

Sugerencias	
✓	Colocar preguntas fáciles, importantes e interesantes al inicio (primero relacionadas con información básica, después clasificación y finalmente identificación)
✓	Agrupar preguntas relacionadas con la misma temática (de manera serial, semántica, de general a específicas)
✓	Iniciar con preguntas de opinión de los participantes
✓	Colocar preguntas de lo específico a lo general sino existe una opinión claramente expresada por los participantes o a falta de un marco de referencia sobre el tema
✓	Acomodar las preguntas sensibles al final
✓	Generar un flujo natural de las preguntas

Nota. Elaboración propia con base en (Taherdoost, 2022).

¿Qué es la validez de un instrumento?

La validez explica qué tan correctamente la información fue recolectada, verifica que los datos cumplan con el objetivo de investigación y que reflejen con precisión el **constructo** (s) a medir. En otras palabras, asegura que un instrumento, como el cuestionario, tenga la capacidad de medir lo que pretende medir (Phillips et al., 2021; Taherdoost, 2016). Resultando una estrategia de análisis preliminar antes de su empleo. Existen distintos tipos de validación, en este documento se describirán cuatro.

La de caso único es definida como el análisis más simple, se aplica a una persona que cumpla con características semejantes a las de la población de estudio, durante este proceso se van respondiendo las preguntas y se retroalimenta sobre la extensión, claridad y cobertura del cuestionario (Carrion et al., 2015).

Otro tipo, es la validez de contenido, la cual tiene como objetivo evaluar el contenido del instrumento por un comité de expertos, que cuenten con trayectoria y reconocimiento en el tema para que sean capaces de analizar los ítems del instrumento, proporcionar información, juicios y valoraciones certeros. Además, que hagan sugerencias de redacción y de contenido. Se suelen utilizar

cuatro categorías para realizar esta evaluación: claridad, coherencia, relevancia y suficiencia (Galicía Alarcón et al., 2017). Algunos de los métodos para analizar este tipo de validez son: análisis factorial de Tucker, Índice de Validez de Lawshe, Índice de Congruencia de Ítem Objetivo de Rovinelli y Hambleton, V de Aiken, entre otros.

La validez de criterio o validez concreta se refiere al nivel en que una medida se asocia con un resultado, evalúa que tanto una medida predice el resultado para otra medida. Por lo que este tipo de validez puede usarse para predecir el rendimiento o comportamiento pasado, presente o futuro (Taherdoost, 2016). Siendo útil cuando se desean hacer inferencias a partir de puntajes respecto a otras variables de interés. Existen dos tipos: **predictiva** y **concurrente**. La primera predice alguna medida del criterio que se va a realizar a futuro con un diseño **prospectivo** (análisis con el fin de explorar o de predecir el futuro), la segunda relaciona las puntuaciones de la prueba con alguna medida del criterio tomada en el mismo momento, esta implica un diseño **transeccional** o **transversal** (que recolecta datos en un único momento en el tiempo) (Argibay, 2006).

La validez de constructo implica que tan efectivamente se ha convertido un constructo en una realidad práctica y útil. Tiene dos elementos: **validez convergente** (en qué nivel la variable latente se distingue de otras) y **validez discriminante** (prueba que dos medidas de constructos que teóricamente deberían estar relacionados lo estén) (Taherdoost, 2016). Para comprobar la validez convergente y discriminante se realiza

un Análisis Factorial Exploratorio (AFE) el cual analiza las intercorrelaciones de un conjunto de datos, permite establecer agrupaciones de ítems correlacionados entre sí que remiten a factores subyacentes no observables y que sirven para medir la validez del **constructo(s)**. Así, permite no sólo evaluar la validez del instrumento, sino también la forma en la que está construido (Argibay, 2006).

¿Qué es la fiabilidad de un instrumento?

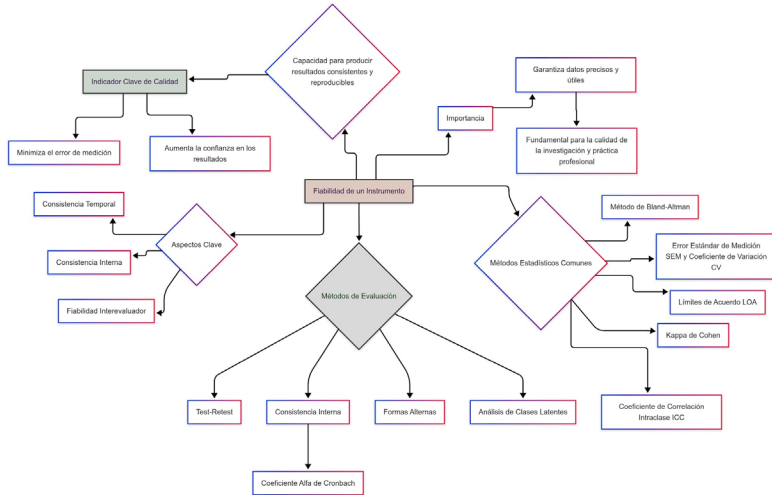
La **fiabilidad** de un instrumento se define como su capacidad para producir resultados consistentes y reproducibles a lo largo del tiempo y bajo distintas condiciones. Es un indicador fundamental de la calidad de un instrumento de medición, trabajando en conjunto con la validez, que asegura que el instrumento mide lo que realmente se propone medir (Kimberlin & Winterstein, 2008; Scholtes et al., 2011; Souza et al., 2017). Es crucial para asegurar que los datos obtenidos sean precisos y útiles tanto para la investigación como para la toma de decisiones. Un instrumento fiable minimiza el error de medición y aumenta la confianza en los resultados obtenidos (Kimberlin & Winterstein, 2008; Scholtes et al., 2011; Souza et al., 2017). En resumen, garantizar la fiabilidad de un instrumento es esencial para la consistencia y precisión de las mediciones, lo que a su vez permite que los resultados sean reproducibles y confiables, elementos fundamentales para la calidad de la investigación y la práctica profesional.

La fiabilidad abarca varios aspectos cruciales. La **consistencia temporal** evalúa la estabilidad de las medidas a lo largo del tiempo, garantizando que los resultados sean uniformes en diferentes momentos. Por otro lado, la **consistencia interna** se refiere a la coherencia de los resultados dentro del mismo instrumento, examinando si sus diferentes partes miden de manera uniforme. Finalmente, la **fiabilidad interevaluador** mide la consistencia de los resultados cuando distintos evaluadores utilizan el mismo instrumento, asegurando que las variaciones no se deban a diferencias entre quienes evalúan (Kimberlin & Winterstein, 2008; Olayinka & Abideen, 2023). Para evaluar la fiabilidad, se utilizan diversos métodos. El **Test-Retest** examina la estabilidad de los

resultados aplicando el mismo instrumento en dos momentos diferentes. La **consistencia interna** se evalúa comúnmente con el coeficiente alfa de Cronbach, que mide la correlación entre los distintos ítems del instrumento. Las **formas alternas** comparan los resultados de diferentes versiones de un mismo instrumento para asegurar su equivalencia. Un método más moderno es el **Análisis de Clases Latentes**, que evalúa la fiabilidad identificando patrones de respuesta consistentes en los datos (Olayinka & Abideen, 2023; Tourangeau et al., 2020; Yusup, 2018).

Existen varios métodos estadísticos para cuantificar la fiabilidad. El **Coefficiente de Correlación Intraclassa (ICC)** es muy utilizado para evaluar la fiabilidad de instrumentos que miden variables continuas, midiendo la consistencia o acuerdo entre diferentes mediciones. El **Método de Bland-Altman** es popular para evaluar el acuerdo entre dos mediciones, permitiendo una visualización gráfica de las diferencias e identificando la **heterocedasticidad** (la varianza de los errores no es igual para todas las variables). El **Error Estándar de Medición (SEM)** y el **Coefficiente de Variación (CV)** se emplean para describir la “fiabilidad absoluta” y son adecuados para comparar la fiabilidad entre distintas herramientas de medición, ayudando a predecir el cambio “real” en las mediciones. Para escalas categóricas, la prueba Kappa de Cohen mide el grado de acuerdo entre evaluadores más allá del azar. Finalmente, los **Límites de Acuerdo (LOA)** proporcionan un rango dentro del cual se espera que caigan la mayoría de las diferencias entre mediciones repetidas, siendo útiles en estudios de comparación de métodos (Atkinson & Nevill, 1998; Park et al., 2018; Raykov & Marcoulides, 2015; Sainani, 2017; Zaki et al., 2012).

Figura 3 Fiabilidad de un instrumento para recolección de datos y métodos de evaluación



Nota. Elaboración propia con base en (Atkinson & Nevill, 1998; Park et al., 2018; Raykov & Marcoulides, 2015; Sainani, 2017; Zaki et al., 2012, 2013)

Relación entre validez y fiabilidad

Siempre debe buscarse que el instrumento de medición, en este caso el cuestionario, mida realmente el constructo (os) determinado y además que proporcione respuestas fiables es decir que sean estables o consistentes (Bolarinwa, 2015). Es importante que el cuestionario sea tanto válido como fiable para que pueda cumplirse con el objetivo de la investigación, esta sea útil, de calidad, permita tomar decisiones adecuadamente y pueda ser comparable en condiciones similares. Las propiedades psicométricas o cualidades (validez y fiabilidad) no son inherentes al instrumento de medición, pueden poseer distintos niveles y están influenciadas por distintos factores, por el constructo (os) que se está midiendo, la población de estudio,

el evaluador e incluso el uso de las puntuaciones generadas del instrumento (Swan et al., 2023). Dichas cualidades al no ser inherentes pueden presentarse una o la otra. Por ejemplo, un termómetro que siempre mide 2°C menos de lo real es fiable pero no válido. Por otro lado, un instrumento puede ser válido, pero no fiable (mide lo correcto, pero con errores inconsistentes) como una escala psicométrica de ansiedad bien diseñada que de resultados muy diferentes al aplicarse dos veces seguidas. Lo ideal es que el instrumento sea válido y fiable, es decir que el termómetro muestre la temperatura correcta (válido) y de consistentemente la misma lectura (fiable) cada vez que sea usado en situaciones comparables (Bourke, 2016; Taherdoost, 2016).

Tabla 2 Recomendaciones y ejemplos para redactar preguntas

Tipo/método	Consejos
Análisis de contenido (validez)	✓ Definir las dimensiones y variables que se pretenden medir, evitar que el instrumento sea diverso y con múltiples dimensiones. ✓ Ser claro en los criterios a evaluar, así como la escala de calificación a emplear. ✓ La extensión de los instrumentos debe ser breve. ✓ Considerar las cargas laborales de los expertos evaluadores, proponer un lapso pertinente para evaluar el instrumento. ✓ Proporcionar plantillas intuitivas y fáciles de responder para agilizar el proceso de recolección de información.
De criterio (validez)	✓ Es fundamental relacionar las puntuaciones del instrumento con el criterio. - Identificar el tipo de validez (concurrente o predictiva) a emplear
De constructo (validez)	✓ Dado que el tamaño de la muestra y el número de ítems influyen en la estabilidad de la estructura factorial, la cantidad de individuos debe ser de al menos cinco por cada ítem del instrumento.
Alfa de Cronbach (fiabilidad)	✓ Este coeficiente es afectado por el número de ítems, al incrementarlos la varianza aumenta. ✓ A mayor número de individuos (muestra) que completen la escala, mayor será la varianza esperada.
Coefficiente Kappa de Cohen (fiabilidad)	✓ Para ejecutar esta prueba, es necesario cumplir las siguientes cinco suposiciones: 1) la(s) variable(s) deben ser medidas categóricamente y ser mutuamente excluyentes, 2) las observaciones deben ser pareadas, 3) cada variable debe contar con los mismos valores de respuesta y la tabulación cruzada debe ser simétrica, 4) los dos codificadores encargados de hacer juicios son independientes para evitar cualquier sesgo, 5) el número de codificadores que hacen juicio sube a dos

Nota. Elaboración propia con base en (Argibay, 2006; Galicia Alarcón et al., 2017; Goyanes, 2024; Oviedo & Campo-Arias, 2005).

¿Cómo mejorar y evaluar las propiedades psicométricas?

Existen distintos factores que influyen en la validez y fiabilidad de un instrumento, por lo que en la tabla 4 se presentan una serie de recomendaciones para su mejora, que faciliten el proceso de análisis y por ende la calidad del cuestionario que ha sido diseñado.

Para evaluar la validez y fiabilidad de un cuestionario, se pueden utilizar diversos programas y

entornos estadísticos. Entre los más utilizados se encuentran R, Python, JASP y JAMOVI, además de hojas de cálculo especializadas y software como SPSS y WINSTEPS. En la tabla 5 se presenta un análisis del software y herramientas, así como su acceso para reconocer las principales opciones que existen para la evaluación de las propiedades psicométricas de un cuestionario.

Tabla 5 Softwares y herramientas para análisis de validez y fiabilidad

Software/ Herramienta	Características Principales	Análisis de Validez	Análisis de Fiabilidad	Características Específicas	Acceso
SPSS	Fácil de usar, manejo de datos faltantes, efectos aleatorios	Modelos mixtos, validez de constructo, efectos fijos y aleatorios, análisis de covarianza	ICC, alfa de Cronbach integrados, análisis de fiabilidad con datos incompletos	Instrucciones específicas disponibles, flexibilidad en diseños complejos	X De pago
AMOS (SPSS)	Ideal para análisis factorial confirmatorio	Análisis factorial confirmatorio avanzado	Índices de fiabilidad del modelo	Modelado de ecuaciones estructurales especializado	X De pago
WINSTEPS	Análisis psicométrico avanzado basado en Rasch	Validez de constructo (Rasch), validez de contenido, análisis de funcionamiento diferencial	Índices de separación, coeficientes de fiabilidad del modelo, fiabilidad de personas e ítems	Modelo de Rasch especializado, mapas de Wright, análisis de ajuste de ítems	X De pago
R (psych, lavaan, irr)	Poderoso, paquetes especializados para psicometría	Pruebas psicométricas avanzadas, validez de contenido, validez de criterio, validez de constructo	Alfa, ICC, consistencia interna, pruebas de estabilidad, equivalencia	Paquetes especializados, análisis psicométricos automatizados, flexibilidad metodológica	✓ Gratis
Python (pingouin, factor_analyzer, semopy)	Alternativa gratuita a SPSS, requiere programación	Módulos para análisis psicométrico, validez de contenido, validez de criterio, validez de constructo	Análisis de fiabilidad automatizado, alfa de Cronbach, pruebas de estabilidad, consistencia interna	Paquetes especializados, análisis psicométricos automatizados, flexibilidad metodológica	✓ Gratis
JASP	Interfaz amigable similar a SPSS, pero gratuita	Análisis factorial, correlaciones, validez de contenido, validez de criterio, validez de constructo	Alfa de Cronbach, análisis de consistencia, pruebas de estabilidad, equivalencia	Paquetes especializados, análisis psicométricos automatizados, flexibilidad metodológica	✓ Gratis
Jamovi	Similar a SPSS, permite análisis de fiabilidad	Validez de constructo básica, validez de contenido, validez de criterio	Análisis de fiabilidad integrado, alfa de Cronbach, pruebas de estabilidad, consistencia interna	Paquetes especializados, análisis psicométricos automatizados, flexibilidad metodológica	✓ Gratis
Excel/Google Sheets (plantillas personalizadas)	Accesible, permite transformaciones detalladas	Regresión, análisis de correlación, regresión lineal, correlación de Pearson	Error intra-sujeto, ICC, alfa (manual), correlación intraclase, desviaciones estándar entre sujetos y pruebas, alfa de Cronbach	Transformaciones logarítmicas para variabilidad proporcional, análisis detallado de consistencia, precisión en mediciones repetidas	✓ Gratis

Nota. Elaboración propia con base en (Hopkins, 2015; Souza et al., 2017; Zhou et al., 2017).

Conclusiones

Para garantizar la calidad de una investigación basada en cuestionarios, es indispensable asegurarse de que los datos recopilados no sean erróneos. Resulta vital que el diseño, desarrollo y evaluación de las propiedades psicométricas del instrumento de recolección de información sean adecuados. Un instrumento bien diseñado no solo garantiza datos confiables y precisos, sino que también minimiza el error de medición y aumenta

la confianza en los resultados obtenidos. La **validez** asegura que el instrumento mida lo que realmente debe medir, mientras que la **fiabilidad** garantiza que los resultados sean consistentes y estables a lo largo del tiempo y bajo distintas condiciones. Además, utilizar herramientas y software estadísticos especializados fortalece los resultados, contribuye significativamente en la mejora de la investigación y en la toma de decisiones.

Conflicto de intereses

Los autores de este manuscrito declaran no tener ningún conflicto de interés.

Declaración de privacidad

Los datos personales facilitados por los autores a RD-ICUAP se usarán exclusivamente para los fines declarados por la misma, no estando disponibles para ningún otro propósito ni proporcionados a terceros.

Descargo de Responsabilidad (INTELIGENCIA ARTIFICIAL). Los autores declaran por la presente que NO se han utilizado tecnologías de IA generativa, tales como modelos de lenguaje grandes (ChatGPT, COPILOT, etc.) y generadores de texto a imagen, durante la redacción o edición de este manuscrito.

Agradecimientos

Los autores agradecen a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) el financiamiento para los estudios de posgrado de los estudiantes José Muñoz Flores, Fatima Andrea Bonilla Ramos y Jazmin Zaragoza Alonso

Referencias

- Argibay, J. C. (2006). Técnicas psicométricas. Cuestiones de validez y confiabilidad. Subjetividad y procesos cognitivos, 8, 15-33 (2006). <https://dspace.uces.edu.ar/jspui/handle/123456789/765>
- Atkinson, G., & Nevill, A. (1998). Statistical Methods For Assessing Measurement Error (Reliability) in Variables Relevant to Sports Medicine. Sports Medicine, 26, 217–238. <https://doi.org/10.2165/00007256-199826040-00002>
- Bolarinwa, O. (2015). Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. Nigerian Postgraduate Medical Journal, 22(4), 195. <https://doi.org/10.4103/1117-1936.173959>
- Bourke, Jane. K. Ann. D. Justin. (2016). 7: VALIDITY & RELIABILITY. En SURVEY & QUESTIONNAIRE DESIGN : Collecting Primary Data to Answer Research Questions. <https://www.ebsco.com/terms-of-use>
- Carrión, C., Soler, M., & Aymerich, M. (2015). Análisis de la Validez de Contenido de un Cuestionario de Evaluación del Aprendizaje Basado en Problemas: Un Enfoque Cualitativo. Formación universitaria, 8(1), 13–22. <https://doi.org/10.4067/S0718-50062015000100003>
- Dalati, S., & Marx Gómez, J. (2018). Survey and questionnaires. En Marx Gómez Jorge & Mouselli Sulaiman (Eds.), Modernizing the Academic Teaching and Research (pp. 175–186). Springer, Cham. https://doi.org/https://doi.org/10.1007/978-3-319-74173-4_10
- DeCastellarnau, A. (2018). A classification of response scale characteristics that affect data quality: a literature review. Quality and Quantity, 52(4), 1523–1559. <https://doi.org/10.1007/s11135-017-0533-4>
- Dueber, D. M., Toland, M. D., Lingat, J. E., Love, A. M. A., Qiu, C., Wu, R., & Brown, A. V. (2022). To Reverse Item Orientation or Not to Reverse Item Orientation, That Is the Question. Assessment, 29(7), 1422–1440. <https://doi.org/10.1177/10731911211017635>
- Elangovan, N., & Sundaravel, E. (2021). Method of preparing a document for survey instrument validation by experts. MethodsX, 8. <https://doi.org/10.1016/j.mex.2021.101326>
- Galicia Alarcón, L. A., Balderrama Trápaga, J. A., Edel Navarro, R., Galicia Alarcón, L. A., Balderrama Trápaga, J. A., & Edel Navarro, R. (2017). Validez de contenido por juicio de expertos: propuesta de una herramienta virtual. Apertura (Guadalajara, Jal.), 9(2), 42–53. <https://doi.org/10.32870/AP.V9N2.993>
- Goyanes, M. (2024). Análisis de contenido en SPSS y KALPHA: Procedimiento para un Análisis Cuantitativo Fiable con la Kappa de Cohen y el Alpha de Krippendorff. Estud. mensaje period., 30(1), 123–140. <https://doi.org/10.5209/espmp.92732>
- Hopkins, W. G. (2015). Spreadsheets for Analysis of Validity and Reliability. SportsScience, 19, 36+. <https://link.gale.com/apps/doc/A562004416/AONE?u=anon-b9d8c65b&ap>
- Kimberlin, C., & Winterstein, A. (2008). Validity and reliability of measurement instruments used in research. American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists, 65(23), 2276–2284. <https://doi.org/10.2146/ajhp070364>
- Krosnick, J. A. (2018a). Improving question design to maximize reliability and validity. En The Palgrave Handbook of Survey Research (pp. 95–101). Springer International Publishing. https://doi.org/10.1007/978-3-319-54395-6_13

- Krosnick, J. A. (2018b). Questionnaire design. En *The Palgrave Handbook of Survey Research* (pp. 439–455). Springer International Publishing. https://doi.org/10.1007/978-3-319-54395-6_53
- Olayinka, S., & Abideen, T. (2023). Reliability of research instruments in management sciences research: an explanatory perspective. *Scientific Papers of Silesian University of Technology. Organization and Management Series*, 166, 46. <https://doi.org/10.29119/1641-3466.2022.166.46>
- Oviedo, H. C., & Campo-Arias, A. (2005). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 34(4), 572–580. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0034-74502005000400009&lng=en&nrm=iso&tng=es
- Park, M., Kang, K., Jang, S., Lee, J., & Chang, S. (2018). Evaluating test-retest reliability in patient-reported outcome measures for older people: A systematic review. *International journal of nursing studies*, 79, 58–69. <https://doi.org/10.1016/j.ijnurstu.2017.11.003>
- Phillips, S. M., Summerbell, C., Hobbs, M., Hesketh, K. R., Saxena, S., Muir, C., & Hillier-Brown, F. C. (2021). A systematic review of the validity, reliability, and feasibility of measurement tools used to assess the physical activity and sedentary behaviour of pre-school aged children. En *International Journal of Behavioral Nutrition and Physical Activity* (Vol. 18, Número 1). BioMed Central Ltd. <https://doi.org/10.1186/s12966-021-01132-9>
- Raykov, T., & Marcoulides, G. (2015). Scale Reliability Evaluation With Heterogeneous Populations. *Educational and Psychological Measurement*, 75, 875–892. <https://doi.org/10.1177/0013164414558587>
- Robinson, M. A. (2018). Using multi-item psychometric scales for research and practice in human resource management. *Human Resource Management*, 57(3), 739–750. <https://doi.org/10.1002/hrm.21852>
- Sainani, K. (2017). Reliability Statistics. *PM&R*, 9. <https://doi.org/10.1016/j.pmrj.2017.05.001>
- Scholtes, V. A., Terwee, C. B., & Poolman, R. W. (2011). What makes a measurement instrument valid and reliable? *Injury*, 42(3), 236–240. <https://doi.org/10.1016/j.injury.2010.11.042>
- Souza, A., Alexandre, N., & Guirardello, E. (2017). Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e serviços de saúde : revista do Sistema Unico de Saúde do Brasil*, 26(3), 649–659. <https://doi.org/10.5123/S1679-49742017000300022>
- Swan, K., Speyer, R., Scharitzer, M., Farneti, D., Brown, T., Woisard, V., & Cordier, R. (2023). Measuring what matters in healthcare: a practical guide to psychometric principles and instrument development. En *Frontiers in Psychology* (Vol. 14). Frontiers Media SA. <https://doi.org/10.3389/fpsyg.2023.1225850>
- Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. En *International Journal of Academic Research in Management (IJARM)* (Vol. 5, Número 3). <https://hal.science/hal-02546799v1>
- Taherdoost, H. (2022). Designing a Questionnaire for a Research Paper: A Comprehensive Guide to Design and Develop an Effective Questionnaire Hamed Taherdoost. *Designing a Questionnaire for a Research Paper: A Comprehensive Guide to Design and Develop an Effective Questionnaire Designing a Questionnaire for a Research Paper: A Comprehensive Guide to Design and Develop an Effective Questionnaire*. *Asian Journal of Managerial Science*, 11(1), 8–16. <https://doi.org/10.51983/ajms-2022.11.1.3087i>
- Tourangeau, R., Sun, H., & Yan, T. (2020). Comparing Methods for Assessing Reliability. *Journal of Survey Statistics and Methodology*, 9(4), 651–673. <https://doi.org/10.1093/issam/smaa018>

- Yusoff, M. S. B., Arifin, W. N., & Hadie, S. N. H. (2021). ABC of questionnaire development and validation for survey research. En *Education in Medicine Journal* (Vol. 13, Número 1, pp. 97–108). Penerbit Universiti Sains Malaysia. <https://doi.org/10.21315/EIMJ2021.13.1.10>
- Yusup, F. (2018). Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Jurnal Tarbiyah : Jurnal Ilmiah Kependidikan*, 7(1). <https://doi.org/10.18592/TARBIYAH.V7I1.2100>
- Zaki, R., Bulgiba, A., Ismail, R., & Ismail, N. (2012). Statistical Methods Used to Test for Agreement of Medical Instruments Measuring Continuous Variables in Method Comparison Studies: A Systematic Review. *PLoS ONE*, 7. <https://doi.org/10.1371/journal.pone.0037908>
- Zaki, R., Bulgiba, A., Nordin, N., & Ismail, N. (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, 803–807. <https://doi.org/10.22038/IJBMS.2013.998>
- Zhou, L., Almutairi, A., Alsaid, N., Warholak, T., & Cooley, J. (2017). Establishing the Validity and Reliability Evidence of Preceptor Assessment of Student Tool. *American Journal of Pharmaceutical Education*, 81. <https://doi.org/10.5688/ajpe5908>