

Validez de contenido por juicio de expertos: Integración cuantitativa y cualitativa en la construcción de instrumentos de medición

Neiber Maldonado-Suárez ^a <https://orcid.org/0000-0003-2053-5232>

Felipe Santoyo-Telles ^b <https://orcid.org/0000-0003-3854-9405>

Universidad de Guadalajara. México.

^a Estudiante del Doctorado en Psicología con Orientación en Calidad de Vida y Salud en el Centro Universitario del Sur de la Universidad de Guadalajara. Cursó la Maestría en Psicología con Orientación en Calidad de Vida y Salud; ambos posgrados pertenecen al PNPC-SNP de CONAHCyT. Av. Enrique Arreola Silva No. 883, Colonia Centro, Código Postal 49000, Ciudad Guzmán, Jalisco, México. C/e: neiber.maldonado@alumnos.udg.mx

^b Doctor en Ciencia y Tecnología. Profesor investigador del Departamento de Ciencias Exactas y Metodologías. Miembro del Claustro Académico de los Posgrados en Psicología con Orientación en Calidad de Vida y Salud, Centro Universitario del Sur, Universidad de Guadalajara, México.

Artículo metodológico. Recibido: 25/03/2024. Revisado: 16/06/2024. Aceptado: 25/06/2024. Publicación avanzada: 29/06/2024. Publicado: 01/07/2024.

Resumen

INTRODUCCIÓN. Algunas expresiones como la inteligencia, la personalidad, las emociones o las percepciones no son tangibles, por lo que se miden mediante constructos compuestos por dimensiones, evaluadas con ítems; esto requiere el desarrollo de instrumentos de medición que permitan inferir la presencia de dichas variables.

MÉTODO. La construcción de estos instrumentos es un proceso complejo, iterativo, con diversas fases de revisión y validación empírica. La validez de contenido por juicio de expertos es una de estas fases; implica someter los ítems al juicio de especialistas para identificar vacíos, ambigüedades o sesgos. Aunque esto asegura la representatividad y relevancia de los ítems, rara vez se informa este procedimiento en la validación de instrumentos.

RESULTADOS. Este artículo sintetiza el procedimiento de validez de contenido por juicio de expertos en la construcción de un instrumento de medición; además, muestra la ejecución del análisis cuantitativo mediante el Coeficiente de Validez de Contenido con apoyo de una macro de Excel, así como la integración cualitativa de las observaciones proporcionadas por los jueces.

CONCLUSIÓN. Se concluye con la importancia de sistematizar e informar sobre las decisiones tomadas en este proceso, a fin de garantizar el rigor y trazabilidad en la construcción de instrumentos de medición.

Palabras clave

instrumento de medición, investigación cualitativa, juicio de expertos, psicometría, validez de contenido.

Referencia recomendada

Maldonado-Suárez, N., y 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.

<https://doi.org/10.1344/reire.46238>

© 2024 Los autores. Este artículo es de acceso abierto sujeto a la licencia Reconocimiento 4.0 Internacional de Creative Commons, la cual permite utilizar, distribuir y reproducir por cualquier medio sin restricciones siempre que se cite adecuadamente la obra original. Para ver una copia de esta licencia, visite

<https://creativecommons.org/licenses/by/4.0/>



Títol (català)

Validesa de contingut per judici d'experts: integració quantitativa i qualitativa en la construcció d'instruments de mesura

Resum

INTRODUCCIÓ. Algunes expressions com la intel·ligència, la personalitat, les emocions o les percepcions no són tangibles, per la qual cosa es mesuren mitjançant constructes compostos per dimensions avaluades amb ítems. Això requereix el desenvolupament d'instruments de mesura que permetin inferir la presència d'aquestes variables.

MÈTODE. La construcció d'aquests instruments és un procés complex, iteratiu, amb diverses fases de revisió i validació empírica. La validesa de contingut per judici d'experts és una d'aquestes fases; implica sotmetre els ítems al judici d'especialistes per identificar-ne buits, ambigüitats o biaixos. Tot i que això assegura la representativitat i la rellevància dels ítems, poques vegades se segueix aquest procediment en la validació d'instruments.

RESULTATS. Aquest article sintetitza el procediment de validesa de contingut segons el parer d'experts en la construcció d'un instrument de mesura; a més, mostra l'execució de l'anàlisi quantitativa mitjançant el coeficient de validesa de contingut, amb el suport d'una macro d'Excel, així com la integració qualitativa de les observacions proporcionades pels jutges.

CONCLUSIÓ. Es conclou amb la importància de sistematitzar i informar sobre les decisions preses en aquest procés per tal de garantir el rigor i la traçabilitat en la construcció d'instruments de mesura.

Paraules clau

instrument de mesura, investigació quantitativa, judici d'experts, psicometria, validesa de contingut.

Title (English)

Content validity by expert judgment: Quantitative and qualitative integration in the construction of measurement instruments

Abstract

INTRODUCTION. Phenomena such as intelligence, personality, emotions, or perceptions are intangible; their measurement requires the use of operational definitions broken down into dimensions evaluated with items. To do so, measurement instruments must be developed that are able to infer the presence of these variables.

METHOD. Constructing these instruments is a complex, iterative process involving various phases of review and empirical validation. Content validity through expert judgment is one of these phases, in which items are subjected to scrutiny by specialists in order to identify gaps, ambiguities, or biases. Although this procedure ensures item representativeness and relevance, it is only rarely reported in instrument validation.

RESULTS. This article summarizes the procedure of content validity by expert judgment in the construction of a measurement instrument; it also describes the execution of the quantitative analysis using the Content Validity Coefficient with the support of an Excel macro, as well as the qualitative integration of the observations provided by the judges.

CONCLUSION. The article stresses the importance of systematizing and reporting the decisions made in this process, in order to guarantee rigour and traceability in the construction of measurement instruments.

Keywords

measurement instrument, quantitative research, expert judgment, psychometrics, content validity.

1. Introducción

La noción de la ciencia es dinámica, pues se construye a partir de las investigaciones que se realizan por las comunidades científicas que se suceden en el tiempo; una primera aproximación, puede ser que se trata de un modo de investigar, es decir, de obtener respuestas a ciertas preguntas (Fullat, 2008). En la actualidad, se le concibe como una actividad sistemática, intensiva e iterativa, cuyo propósito es elaborar explicaciones basadas en teorías y modelos, sobre los principios que rigen los sistemas que constituyen la realidad. Las teorías y modelos científicos no solo permiten representar hechos, sino que hacen posible la predicción de su comportamiento; el conocimiento de los principios que rigen la realidad permite crear soluciones prácticas e innovadoras para desafíos complejos, como el tratamiento de enfermedades, el desarrollo de nuevas tecnologías, así como la resolución de problemas críticos (Bartlett *et al.*, 2023; Danchin, 2023).

La ciencia moderna finca sus fundamentos en el positivismo; esta corriente sostiene que la realidad se rige por leyes aprehensibles para el ser humano. Bajo esta visión, la tarea del científico es descubrir dichas leyes para explicar los hechos, hacer predicciones y modificar algunos aspectos de la realidad. Además, se espera que quien investiga sea objetivo respecto al objeto que estudia, con el propósito de asegurar que el conocimiento refleje el funcionamiento de la realidad (Fullat, 2008; Park *et al.*, 2020).

Las principales críticas al positivismo recaen en su perspectiva reduccionista, la búsqueda de objetividad, el énfasis en la cuantificación, predicción y control, así como en su falta de perspectiva histórica y social (Ruiz Taborda *et al.*, 2021). Como respuesta a estos planteamientos, surgió el post-positivismo que sostiene que la realidad no puede ser aprehendida en su totalidad, sino que se conoce de forma imperfecta a partir de la interacción entre el objeto de estudio y el sujeto cognoscente (Ramos, 2015). Estas dos visiones han marcado la clasificación del conocimiento, en función de su metodología y objeto de estudio.

Durante la investigación científica, es fundamental recabar datos sobre los hechos en los que se tiene interés; desde la perspectiva cuantitativa, a este proceso se le conoce como medición y consiste en la asignación de valores para representar las propiedades de los objetos y eventos del mundo (Coolican, 2004). Cuando se pretende medir un hecho observable, que se manifiesta de forma física y material, existen instrumentos, generalmente de tipo mecánico, para obtener mediciones confiables y comparables (Galindo-Domínguez, 2020).

Existen, en cambio, otros atributos de la realidad que, por su naturaleza abstracta, no se manifiestan de manera tangible, como sucede con la inteligencia, la personalidad, las emociones o las percepciones, lo que representan un desafío para su medición. Para solventarlo, en psicología como en otras disciplinas como en educación y ciencias sociales, a las variables de interés se les trata como constructos, toda vez que no son directamente observables y, por lo tanto, no se pueden medir en primera instancia. Para abordar esta dificultad, es necesario establecer definiciones operacionales, es decir, traducir el concepto abstracto a variables manifiestas. Estas variables se agrupan en dimensiones que, a su vez, se subdividen en indicadores específicos que son evaluados mediante los ítems que componen un instrumento (Coolican, 2004; Kerlinger & Lee, 2002).

La complejidad de los constructos, requiere el desarrollo y validación de instrumentos de medición (escalas, test estandarizados y otros métodos de cuantificación) para determinar en qué sujetos y en qué medida, se presenta la característica que se pretende evaluar; se trata de un proceso caracterizado por diversas fases de revisión y validación empírica, con el propósito de garantizar mediciones precisas en la población de interés (Cervantes Luna *et al.*, 2017; Coolican, 2004; Galindo-Domínguez, 2020; Kerlinger & Lee, 2002).

Por lo anterior, someter los ítems de un instrumento en desarrollo a la evaluación de un panel de especialistas permite detectar vacíos conceptuales, ambigüedades semánticas, construcciones gramaticales inadecuadas, sesgos potenciales, elementos irrelevantes o insuficientes, así como hacer ajustes para garantizar que los ítems evalúen apropiadamente el constructo de interés en la población objetivo. La integración cuantitativa, mediante el análisis estadístico de la concordancia de las opiniones del panel de expertos, permite identificar los ítems que requieren modificaciones; mientras que la integración cualitativa recoge observaciones y sugerencias detalladas de los expertos para ampliar, eliminar o modificar los ítems. Estos procedimientos actúan como filtros que permiten optimizar la representatividad, relevancia, coherencia y comprensibilidad de los elementos que componen el instrumento, fortaleciendo la evidencia sobre su validez, antes de realizar otras fases en el proceso de validación, como los análisis psicométricos (Galicia Alarcón *et al.*, 2017; Pedrosa *et al.*, 2014; Urrutia Egaña *et al.*, 2015).

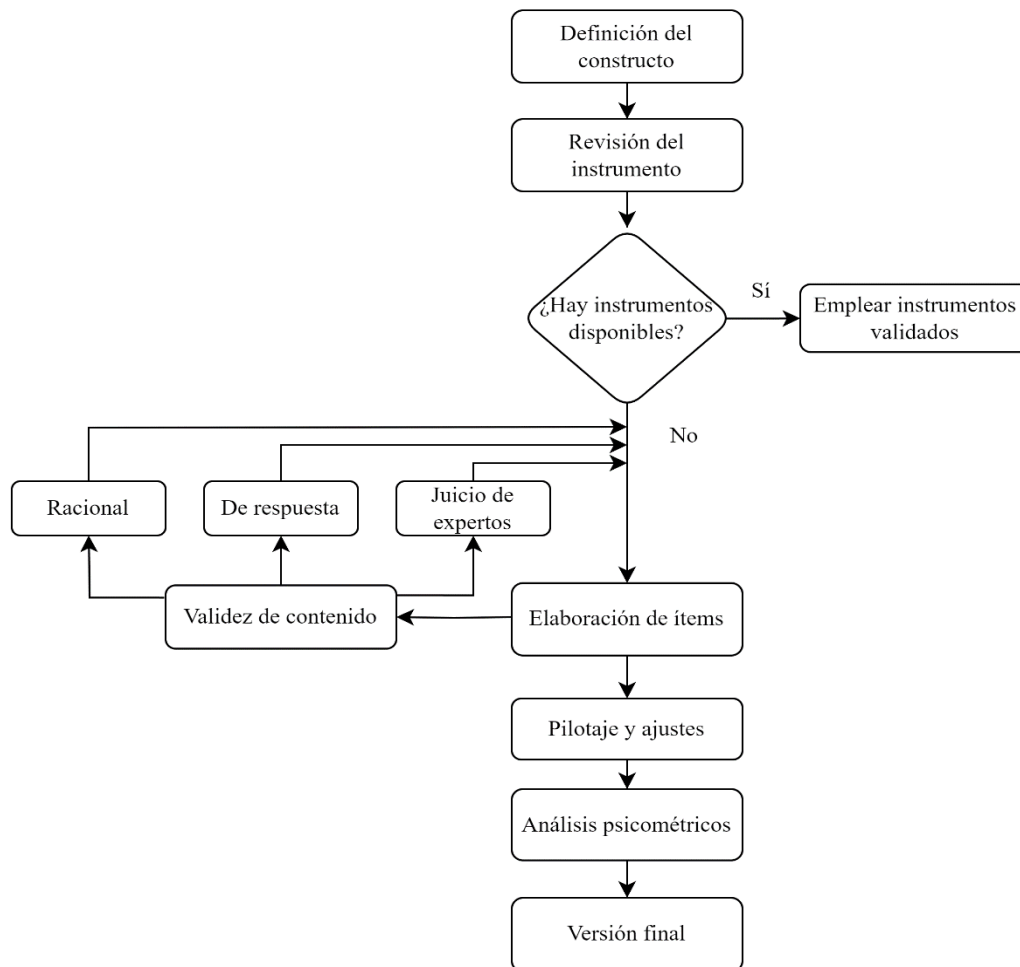
Con base en lo anterior, se efectuó una búsqueda en la base de datos PubMed en la que se introdujeron los términos: “(((Instrument validation OR validación de instrumentos) AND (Research OR Investigación)) AND (Psychology OR Psicología)) AND (Mexico)”, considerando un periodo de 10 años, comprendido de 2014 a 2024; de esta búsqueda, se identificaron 143 trabajos de investigación en los que se reporta algún procedimiento de validación de los instrumentos de medición, de los cuales, únicamente tres estudios (Aizpitarre *et al.*, 2017; Juárez-García *et al.*, 2020; Licona *et al.*, 2014) refirieron haber realizado el proceso de validación por juicio de expertos, sin que se mencionara el análisis cuantitativo de dicho procedimiento.

Considerando lo expuesto, el presente artículo se plantea con el propósito de sintetizar el procedimiento para realizar la validez de contenido por juicio de expertos en la construcción de instrumentos de medición; además, se explica la ejecución del análisis cuantitativo mediante el Coeficiente de Validez de Contenido propuesto por Hernández-Nieto (2002) con apoyo de una macro de Excel diseñada para este propósito, así como la integración cualitativa de las observaciones proporcionadas por el panel de expertos. Para lograrlo, en primer lugar, se presentarán generalidades sobre el procedimiento para diseñar instrumentos de medición, destacando la importancia del juicio de expertos en este proceso. Enseguida, se explorarán aspectos conceptuales relacionados con el juicio de expertos, para introducir el uso de una rúbrica diseñada para orientar al panel de expertos en sus evaluaciones. Posteriormente, se abordará la integración cuantitativa de las evaluaciones emitidas por el panel de expertos a través del Coeficiente de Validez de Contenido de Hernández-Nieto (2002), junto con la aplicación práctica mediante una macro en Excel. Por último, se detallará el proceso de integración de las observaciones cualitativas emitidas por los jueces, enfatizando su papel en la mejora del instrumento de medición.

2. Construcción de instrumentos de medición

La construcción y validación de un instrumento de medición es un proceso que implica múltiples pasos, para asegurar que las mediciones del constructo elegido en la población de interés sean precisas y confiables. Este proceso puede observarse en la *Figura 1* (American Educational Research Association *et al.*, 2018; International Test Commission, 2017).

Figura 1
Construcción de los instrumentos de medición



Fuente: adaptado de American Educational Research Association *et al.* (2018) e International Test Commission (2017).

2.1. Definición del constructo

El primer paso es identificar la característica que se desea medir; se sugiere realizar una revisión de la literatura en bases de datos especializadas, así como solicitar la asesoría de expertos para definir con claridad el nombre del constructo de interés, desde la disciplina en la que se lleva a cabo el estudio (Coolican, 2004; Gaxiola Romero & Manríquez Betanzos, 2021; Kerlinger & Lee, 2002).

2.2. Revisión de instrumentos existentes

Antes de construir un instrumento, se sugiere realizar una revisión de la literatura, que puede seguir los lineamientos para revisiones sistemáticas de la declaración PRISMA (Page *et al.*, 2021), con el fin de identificar si existen instrumentos que midan el constructo de interés, para evitar la duplicidad y aprovechar las fortalezas de las herramientas validadas. Si se encuentran instrumentos que puedan ser de utilidad para los propósitos del estudio, se sugiere emplearlos y descartar la construcción de un nuevo instrumento.

2.3. Desarrollo del instrumento

Cuando no se encuentren disponibles instrumentos que evalúen el constructo de interés, se recomienda iniciar el proceso de construcción del instrumento, que comienza con la elaboración de los ítems. Estos ítems pueden ser estímulos, preguntas o enunciados. Los estímulos son elementos visuales que se utilizan en pruebas perceptuales o neuropsicológicas para activar funciones específicas que se desea evaluar. Por otra parte, algunos instrumentos presentan sus ítems mediante preguntas cerradas con múltiples opciones de respuestas, mientras que otros, presentan sus reactivos en enunciados, que son afirmaciones o proposiciones que la persona evalúa en una escala tipo Likert, para indicar su grado de acuerdo o desacuerdo. Es importante garantizar que la redacción de los ítems sea clara, comprensible y pertinente para el constructo que se pretende medir.

2.4. Validez de Contenido

Este procedimiento está estrechamente vinculado al paso anterior, pues no es posible continuar con la construcción del instrumento sin verificar la validez de contenido. La validez de contenido busca determinar si los ítems propuestos para un instrumento recogen todas las partes que pudieran conformar el constructo psicológico que se pretende medir (Galicia Alarcón *et al.*, 2017; Galindo-Domínguez, 2020; Mousazadeh *et al.*, 2017). Existen tres formas de conocer la validez de contenido: a) validez racional, b) validez de respuesta y c) juicio de expertos.

La validez racional implica realizar una revisión de la literatura del constructo de interés para identificar los elementos que deben componer el instrumento. En ocasiones, no existe suficiente literatura sobre el constructo en cuestión, por lo que se sugiere realizar el proceso de validez de respuesta, que implica hacer entrevistas cualitativas a la población de interés, a fin de identificar los elementos a considerar para integrar el instrumento de medición. Aunque la validez racional puede proporcionar suficiente información para prescindir del proceso de validez de respuesta, optar por ambos métodos puede ocasionar una representación más completa de los ítems que componen el instrumento. En este contexto, los ítems son las unidades que conforman un instrumento, diseñados para recolectar información sobre aspectos específicos del constructo que se está estudiando; la suma de las puntuaciones de los ítems proporciona una medida del constructo o de sus componentes, cuando se trata de un constructo multidimensional (Escobar-Pérez & Cuervo-Martínez, 2008; Galicia Alarcón *et al.*, 2017).

La tercera forma de realizar la validez de contenido es mediante el juicio de expertos; implica conformar un grupo a quienes se les designa como personas expertas. Se les elige debido a que su trayectoria, ya sea en investigación o en el ejercicio profesional, evidencia que tienen conocimientos tanto de la disciplina como del constructo que se pretende evaluar. Así pues, si se está construyendo un instrumento para evaluar la inteligencia infantil, las personas expertas pudieran ser investigadoras en psicología o educación dedicadas a la infancia, con especialidad en psicología cognitiva o psicología del desarrollo. Aunque no existe consenso sobre los indicadores objetivos para definir quién es experto, autores como Perroca (2011) sugieren criterios basados en la temporalidad, estableciendo un mínimo de cinco años de experiencia para conformar el grupo que evaluará el instrumento. Dos aspectos fundamentales a considerar es que la conformación del grupo de expertos dependerá de las especificidades del constructo en el que se tenga interés, además de que se recomienda documentar cuidadosamente el proceso de evaluación por juicio de expertos (Galindo-Domínguez, 2020; Urrutia Egaña *et al.*, 2015).

Se sugiere que quienes participen como personas expertas sean independientes entre ellas, así como en relación con el equipo que está construyendo el instrumento, con el fin de garantizar la mayor objetividad posible. Este grupo de personas expertas evalúa las propiedades de los ítems que componen el instrumento para realizar las modificaciones oportunas. Los criterios para evaluar las propiedades de los ítems pueden definirse según los

aspectos que desee observar el equipo que está construyendo el instrumento. Por ejemplo, si se desea que el instrumento evalúe un constructo con el menor número de ítems posible, un criterio a considerar sería la extensión. Aunque existe libertad en la selección de parámetros para evaluar, el consenso sugiere hacerlo a partir de criterios como suficiencia, claridad, coherencia y relevancia (Escobar-Pérez & Cuervo-Martínez, 2008; Galindo-Domínguez, 2020).

2.5. Pilotaje y ajustes

Una vez que se haya conseguido una versión preliminar del instrumento a partir del juicio de expertos, se administra esta versión a una muestra pequeña, usualmente conformada por 10 o 15 personas, siguiendo el criterio de saturación teórica que implica suspender la aplicación cuando las respuestas no aportan nueva información (Urpí-Fernández *et al.*, 2020), para evaluar la comprensión, aceptabilidad y aplicabilidad del cuestionario, así como el tiempo requerido para su resolución. Este procedimiento se realiza con un método cualitativo denominado entrevistas cognitivas, desarrollado en la década de 1980 por psicólogos y metodólogos; este método permite identificar y corregir problemas en la redacción de los ítems, a partir del análisis de las respuestas de los sujetos en los que se probó el instrumento (Caicedo Cavagnis & Zalazar-Jaime, 2018; Willis, 1999). El nombre de este método se debe a que se fundamenta en la teoría cognitiva, principalmente en el modelo de Tourangeau (1984), que se centra en analizar los procesos cognitivos que las personas realizan para interpretar, recordar, tomar decisiones y responder a las preguntas.

2.6. Análisis psicométricos

En este paso, es necesario identificar una muestra para llevar a cabo los análisis de fiabilidad y validez. Una duda recurrente en este paso es calcular el tamaño de muestra necesario para hacer los análisis psicométricos. Algunas sugerencias plantean lo siguiente:

- Cantidad absoluta de casos: se sugiere un valor mínimo de 200 casos (Vargas Halabí & Mora-Esquivel, 2017).
- Comrey & Lee (1992) sugieren que, aunque la decisión sobre el tamaño de la muestra es del equipo investigador, muestras de tamaño 50 proporcionan confiabilidad inadecuada en los coeficientes de correlación; mientras que muestras de 1000 o más son consideradas adecuadas para los propósitos del análisis factorial. La idoneidad del tamaño de la muestra puede evaluarse en la siguiente escala: 50, muy pobre; 100, pobre; 200, adecuado; 300, buena; 500, muy buena, y 1000 o más, excelente.
- Nunnally (1967) recomienda al menos 10 participantes por cada ítem del instrumento.

Los análisis psicométricos permiten decidir si se eliminan o modifican ciertos ítems para optimizar la calidad y estructura del instrumento. Después de las entrevistas cognitivas, el instrumento con las modificaciones realizadas se prueba en una muestra inicial obtenida de acuerdo con los criterios que se señalaron previamente en este apartado. El objetivo de esta fase es analizar la estructura factorial del instrumento; en esta parte, los ítems se eliminan tomando en cuenta a) su capacidad para discriminar entre diferentes grupos; b) el análisis de consistencia interna con respecto a la puntuación total, de modo que los ítems con valores menores a 0,30 tienden a eliminarse; c) análisis factorial exploratorio, se eliminan aquellos ítems que: 1) tengan cargas similares en dos o más factores; 2) ítems con cargas inferiores a 0,30; 3) factores con menos de tres ítems (Field, 2018; Lloret-Segura *et al.*, 2014).

Una vez identificada la estructura factorial del instrumento, es decir, después de determinar si el constructo está compuesto por dimensiones específicas, se procede con el análisis factorial confirmatorio. Los criterios

empleados en esta evaluación son: a) Chi cuadrado dividido por grados de libertad (χ^2/gl): se considera un ajuste aceptable si $\chi^2/\text{gl} < 3$ (Carmines & McIver, 1981); b) *Comparative Fit Index* (Índice de ajuste comparativo, CFI): se considera adecuado si $\text{CFI} > 0,90$ (McDonald & Marsh, 1990); c) *Tucker-Lewis Index* (Índice de Tucker-Lewis, TLI): se considera adecuado si $\text{TLI} > 0,90$ (Bollen, 1989); d) *Root Mean Square Error of Aproximation* (Raíz del error cuadrático medio, RMSEA): se considera un ajuste razonable si $\text{RMSEA} < 0,08$, rechazando si $\text{RMSEA} > 0,1$ (Browne & Cudeck, 1992). Además, se efectúa el análisis de fiabilidad, ya sea mediante alfa de Cronbach u omega de McDonald. Se evalúa la carga factorial de los ítems, y si se considera pertinente, se eliminan aquellos inferiores a 0,30, siempre que no afecte la estructura del instrumento.

2.7. Versión final del instrumento

Con base en los resultados de las pruebas psicométricas, se elabora la versión final, que incluye instrucciones claras, la guía de aplicación, así como la declaración del consentimiento de confidencialidad por parte del equipo de investigación (CIOMS, 2017; Urpí-Fernández *et al.*, 2020).

2.8. Documentación y diseminación de Resultados

Por último, es preciso reportar las propiedades psicométricas detallar las instrucciones de uso; lo anterior, para diseminar la información junto con el instrumento en una revista científica, con el fin de que investigadores y profesionales interesados en el tema, empleen el instrumento de manera informada.

3. Validez de contenido por juicio de expertos

Una de las etapas importantes en la construcción de instrumentos de medición es la validez de contenido por juicio de expertos. Esta fase implica someter los ítems del instrumento a la evaluación de personas expertas, para determinar si representan todas las dimensiones del constructo que se pretende medir (Escobar-Pérez & Cuervo-Martínez, 2008; Galicia Alarcón *et al.*, 2017; Pedrosa *et al.*, 2014). A partir de la propuesta de Escobar-Pérez y Cuervo-Martínez (2008) y Galicia Alarcón *et al.* (2017) se plantea el procedimiento que se presenta en los párrafos subsecuentes para realizar el proceso de validez de contenido por juicio de expertos.

3.1. Definición de perfiles para constituir el panel de expertos

El primer paso es definir el perfil de las personas que serán elegidas expertas. Se sugiere convocar a personas con experiencia en la temática, que se dediquen a la investigación científica, además de que tengan imparcialidad respecto al equipo investigador y al instrumento. Con esto, se busca maximizar el aporte de conocimiento relevante y minimizar posibles sesgos durante la evaluación.

3.2. Cantidad de integrantes del panel de expertos

Aunque no existe consenso sobre el número de integrantes para el panel de expertos, se sugieren algunos criterios al respecto. Por un lado, es importante valorar las posibilidades de contactar y comunicarse fluidamente con los potenciales expertos, así como su disponibilidad de tiempo, pues contar con panelistas altamente calificados, pero con poco tiempo o motivación, puede ralentizar el proceso. Además, es necesario determinar si existe un número suficiente de profesionales que cumplan con el perfil establecido (Ortega Mohedano, 2008; Urrutia Egaña *et al.*, 2015). Aunque hay diversas opiniones sobre la cantidad ideal, Hernández-Nieto (2002, 2011)

plantea que es posible realizar este procedimiento y ejecutar el análisis cuantitativo mediante el Coeficiente de Validez de Contenido con un grupo de entre tres y cinco jueces.

3.3. Juicio de expertos

Una vez identificado el panel de expertos, se les proporciona el instrumento a evaluar, así como una guía para realizar el jueceo. Escobar-Pérez & Cuervo-Martínez (2008) sugieren cuatro criterios para emitir las opiniones: coherencia, relevancia, claridad y suficiencia. La descripción de estos criterios puede observarse en la *Tabla 1*.

Tabla 1

Criterios para hacer la evaluación por juicio de expertos de un instrumento de medición

Criterio	Indicador
Suficiencia: contiene la cantidad necesaria de ítems por dimensión para su completa evaluación.	<ol style="list-style-type: none"> 1. Los ítems no son suficientes para representar adecuadamente la dimensión. 2. Los ítems miden algún aspecto de la dimensión, pero no corresponden con la dimensión total. 3. Se requiere incrementar el número de ítems para evaluar todos los aspectos de la dimensión. 4. El número de ítems es adecuado y suficiente para medir todos los elementos de la dimensión del constructo.
Claridad: el ítem se comprende fácilmente, debido a la adecuación de su estructura y contenido para la población objetivo.	<ol style="list-style-type: none"> 1. El ítem no es claro. 2. El ítem requiere modificaciones en el uso de las palabras de acuerdo con su significado o bien, por la ordenación de estas. 3. Se requiere modificaciones específicas en algunos de los términos del ítem (es recomendable escribir las recomendaciones específicas en la rúbrica diseñada para la evaluación del instrumento). 4. El ítem se comprende fácilmente, su estructura y contenido son adecuados para la población objetivo
Coherencia: el ítem refleja coherencia conceptual respecto a la dimensión al cual se asocia dentro del constructo	<ol style="list-style-type: none"> 1. El ítem no tiene relación lógica con la dimensión. 2. El ítem tiene una relación débil con la dimensión 3. El ítem tiene una relación moderada con la dimensión que está midiendo. 4. El ítem tiene una relación conceptual fuerte y congruente con la dimensión que busca representar
Relevancia: el ítem aporta información relevante sobre un aspecto necesario de la dimensión que se pretende medir, por lo que debe mantenerse en la versión final del instrumento	<ol style="list-style-type: none"> 1. El ítem puede ser eliminado sin que se vea afectada la medición de la dimensión. 2. El ítem tiene poca relevancia para la dimensión, de modo que su contenido podría estar ya cubierto en otro ítem 3. El ítem tiene relevancia moderada en relación con la representatividad de la dimensión 4. El ítem contiene elementos altamente relevantes para representar adecuadamente la dimensión

Fuente: Adaptado de Escobar-Pérez y Cuervo-Martínez (2008).

Para registrar el juicio del panel de expertos, se diseñó la rúbrica que se muestra en la *Tabla 2*; en la columna "Dimensión" se enumeran las dimensiones en las que se divide el instrumento; en la columna "No. ítem", es posible agregar tantos números como ítems tenga el instrumento. El juez deberá identificar el ítem que está evaluando y asignarle un valor de 1 a 4 en una escala tipo Likert, de acuerdo con la descripción de los criterios señalados en la *Tabla 1*. La rúbrica también permite hacer una evaluación cualitativa en dos secciones: en la

columna “Observaciones” para comentarios específicos de redacción por ítem, y un espacio al pie de la rúbrica para sugerir dimensiones adicionales, si el evaluador lo considera necesario.

Una vez que los jueces evalúan los ítems, el equipo investigador suma las puntuaciones otorgadas por cada experto. Las puntuaciones pueden ir de 4 puntos (si todos los criterios reciben 1 punto) a una puntuación máxima de 16 puntos (si todos los criterios reciben 4 puntos), según la escala de valoración de la rúbrica (*Tabla 2*). Estas sumas se registran en la columna “Suma por cada ítem”.

Tabla 2

Rúbrica para realizar el procedimiento de validez de contenido por juicio de expertos

Dimensión	No. ítem	Califique cada ítem del 1 al 4, de acuerdo con la descripción de la Tabla 1				Observaciones	Suma por cada ítem
		Suficiencia	Coherencia	Relevancia	Claridad	(Escriba libremente sus observaciones al ítem)	
1.ª dimensión	1						
	2						
	3						
	4						
2.ª dimensión	5						
	6						
	7						
	8						
3.ª dimensión	9						
	10						
	11						
	12						
4.ª dimensión	13						
	14						
	15						
	16						
5.ª dimensión	17						
	18						
	19						
	20						
Si considera que hace falta incluir alguna dimensión, por favor describa cómo nombraría la dimensión y qué ítems considera que debería incluir: _____							

Fuente: adaptado de Escobar-Pérez y Cuervo-Martínez, 2008.

3.4. Análisis cuantitativo del juicio de expertos

Es indispensable que el panel de expertos emita valoraciones cuantitativas y no solo cualitativas sobre cada ítem sometido a valoración, que luego se analizan mediante técnicas estadísticas, lo que permite determinar qué ítems son adecuados y cuáles deben ser modificados o eliminados para la versión final del instrumento. Pedrosa *et al.*

(2014) identificaron diversos métodos estadísticos para este análisis en el proceso de validez de contenido (Tabla 3).

Tabla 3

Métodos estadísticos para el análisis cuantitativo del juicio de expertos en el proceso de validación de un instrumento

Autor y año	Método	Descripción
Cohen (1960)	Kappa de Cohen	<p>Compara la tasa de concordancia observada con la tasa de concordancia esperada por el azar, brindando un índice cuantitativo del nivel de acuerdo entre los jueces. Sus valores van de 0 a 1. Su expresión matemática es la siguiente: $\kappa = \frac{(Po - Pe)}{(1 - Pe)}$</p> <p>La proporción de concordancia observada (Po) se cuenta el número de ítems en que los jueces están de acuerdo y se divide entre el total de casos. La proporción de concordancia esperada por azar (Pe) se calcula por juez según cómo distribuyó sus calificaciones. Se obtiene la proporción de casos que calificó "adecuados" y "no adecuados" para cada evaluador. Luego se multiplican estas proporciones entre los distintos jueces, y se suman los productos resultantes.</p>
Tucker (1961)	Método basado en el Análisis Factorial	<p>Se basa en el análisis factorial de las puntuaciones de relevancia dadas por los expertos a cada ítem; permite obtener dos factores:</p> <ol style="list-style-type: none"> 1.- Adecuación muestral de los ítems, pues se considera al test como una muestra representativa de la variable de interés. 2.- Diferencias en las puntuaciones de los expertos al evaluar los ítems.
Fleiss (1971)	Kappa de Fleiss	<p>Extiende el uso de kappa a más de dos evaluadores para medir la concordancia de acuerdo entre más de dos evaluadores. Ajusta la concordancia observada entre múltiples evaluadores al considerar la proporción de acuerdos que podrían haber ocurrido por azar.</p>
Lawshe (1975)	Índice de Validez de Contenido (IVC)	<ol style="list-style-type: none"> 1.- Los jueces, asignan puntuaciones a cada ítem a partir de tres posibilidades: a) que sea esencial, b) que resulte útil pero prescindible o c) que sea innecesario. 2.- Mediante la Razón de Validez de Contenido se determinan los ítems del instrumento que deben mantenerse en la versión final 3.- Finalmente se calcula el Índice de Validez de Contenido que es un promedio de los ítems seleccionado en el paso anterior.
Rovinelli y Hambleton (1977)	Índice de congruencia ítem-objetivo	<ol style="list-style-type: none"> 1.- Cada juez valora el ítem con +1 o -1 si el ítem mide o no el objetivo que se plantea. 2.- La selección de los ítems que constituirán el instrumento, se realiza en función del grado de acuerdo esperado por el investigador en función de los objetivos que haya planteado para el instrumento.
Hambleton (1984)	Índice de congruencia	<ol style="list-style-type: none"> 1.- Una primera versión, permite comparar el grado en que un ítem se relaciona con su dimensión. 2.- La segunda versión del Índice de congruencia, permite también evaluar el ajuste del cada ítem en función del instrumento total.
Aiken (1980)	V de Aiken	<ol style="list-style-type: none"> 1.- Permite evaluar la relevancia de cada ítem respecto a su constructo. 2.- Además, toma en cuenta el número de personas expertas participantes. 3.- A partir de estos datos, se establece el grado de acuerdo basado en la distribución normal, para obtener una probabilidad asociada a cada ítem.

Autor y año	Método	Descripción
Sireci y Geisinger (1992)	Escalamiento multidimensional y análisis de clúster	<p>Este procedimiento busca evaluar similitud de los ítems basándose en el escalamiento multidimensional y el análisis de <i>clúster</i>. Con esto, se pretende solucionar el sesgo de las personas expertas cuando conocen las especificaciones del contenido que se pretende valorar.</p> <ol style="list-style-type: none"> 1.- Se presentan los ítems al panel de expertos para que los asocien con base en su similitud. 2.- Los ítems similares serán agrupados formando un <i>clúster</i> y se encontrarán muy próximos entre sí al momento de realizar el escalamiento multidimensional.
Levine <i>et al.</i> (1997)	Método de Capacidades Mínimas	<p>Este método se enfoca en la selección de personal.</p> <ol style="list-style-type: none"> 1.- Se establece un perfil con las características que cada trabajador debe poseer en función del puesto al que está aspirando. 2.- El panel de experto define el nivel de habilidad mínimo que debe poseer el aspirante al puesto. 3.- Finalmente, las personas expertas evalúan la dificultad para alcanzar el nivel mínimo que se requiere para cada tarea, y el nivel que tiene cada aspirante en función con las tareas establecidas.
Fitch <i>et al.</i> , (2001)	Rango Interpercentil Ajustado a la Simetría	<ol style="list-style-type: none"> 1.- En una escala tipo Likert de 9 puntos, los expertos emiten su juicio respecto a la adecuación y relevancia de los ítems. 2.- Los ítems que se mantienen son aquellos que tienen una mediana superior a siete. 3.- Por último, se calcula el rango interpercentil para determinar el acuerdo entre los distintos expertos respecto al ítem. Se mantienen aquellos que se encuentren entre el 30 % y el 70 %.
Hernández-Nieto (2002)	Coeficiente de Validez de Contenido	<p>Este coeficiente se distingue porque permite valorar el acuerdo de jueces a partir de la participación de entre 3 y 5 personas expertas. El autor recomienda mantener aquellos ítems con un CVC superior a 0,80.</p>
Rubio <i>et al.</i> (2003)	Índice de Validez Factorial	<p>Este método calcula tres índices: la Fiabilidad de Acuerdo Interjueces (IRA); el Índice de Validez de Contenido (IVC) y el Índice de Validez Factorial (FVI).</p> <ol style="list-style-type: none"> 1.- El IRA estima la confiabilidad inter-jueces a partir de la representatividad y claridad de los ítems. Se basa en una escala Likert de 4 puntos que luego se dicotomiza, para identificar los ítems adecuados. El IRA total resulta de dividir los ítems apropiados entre el total de ítems del instrumento. 2.- El IVC se calcula a partir del método propuesto por Lawshe (1975). 3.- El FVI permite evaluar cuánto los expertos asocian cada ítem con los constructos a medir; para calcularlo se divide el número de jueces que vinculan adecuadamente el ítem con su dimensión teórica entre el total de jueces.

Fuente: adaptado de Pedrosa *et al.* (2014)

4. Coeficiente de validez de contenido

Uno de los propósitos de este artículo es explicar la integración cuantitativa y cualitativa de las observaciones, mediante el Coeficiente de Validez de Contenido de Hernández-Nieto (2002, 2011). La fortaleza de este enfoque radica en su posibilidad para aplicarse en situaciones donde no es posible contar con un amplio panel de expertos, por lo que se sugiere realizarlo con un grupo de entre tres y cinco expertos, aunque también es adaptable para paneles con un mayor número de integrantes. El análisis cuantitativo permite obtener dos indicadores: a) El

Coefficiente de Validez de Contenido de cada Ítem (CVC_{ic}) y b) el Coeficiente de Validez de Contenido Total (CVC_t).

El Coeficiente de Validez de Contenido de cada Ítem (CVC_{ic}), “se define como la proporción relativa con respecto del valor máximo de la escala del promedio de los puntajes entre jueces por cada ítem, corregido por concordancia aleatoria (Pe_i)” (Hernández-Nieto, 2011, p. 102). La fórmula es:

$$CVC_{ic} = \left[\frac{\sum xi/j}{Vmx} \right] - Pe_i$$

1. En primer lugar, es necesario obtener la media de las puntuaciones asignadas a cada ítem por los expertos que participaron en la evaluación ($\sum xi/j$). Esta operación se realiza mediante la suma de puntuaciones recabadas en la columna “Puntuaciones por cada ítem” (Tabla 2), dividido entre el total de expertos participantes.
2. Posteriormente, se obtiene el CVC_i (Coeficiente de validez de Contenido sin corrección por concordancia aleatoria):

$$CVC_i = \frac{\sum xi/j}{vmax}, \text{ donde:}$$

- $\sum xi/j$: media de las puntuaciones de los jueces participantes para cada ítem
 - $Vmax$: puntuación máxima que el ítem podría alcanzar.
3. Se parte de la premisa que el CVC_i tiene un error, debido a que, pese a que cada juez evaluó cada ítem de forma independiente, existe la probabilidad de que dos o más jueces asignen puntajes de forma aleatoria; por ello, es preciso calcular el error por concordancia aleatoria (Pe_i) mediante la ecuación:

$$Pe_i = \left(\frac{1}{j} \right)^j, \text{ donde:}$$

- Pe_i : error asignado a cada ítem
 - j : número de expertos participantes
4. Finalmente se calcula el Coeficiente de Validez de Contenido de cada Ítem corregido por concordancia aleatoria CVC_{ic} , de la siguiente forma: $CVC_{ic} = CVC_i - Pe_i$. La puntuación obtenida, es la validez de contenido y la concordancia entre jueces para cada ítem del instrumento.
 5. El cálculo del Coeficiente de Validez de Contenido Total (CVC_t), se obtiene promediando las puntuaciones del CVC_{ic} entre el número de jueces participantes (N). La fórmula es:

$$CVC_t = \frac{\sum CVC_{ic}}{N} = \left[\left[\frac{\sum xi/j}{Vmx} \right] - Pe_i \right] - (1/N)$$

Para realizar el cálculo del Coeficiente de Validez de Contenido, se desarrolló una macro en el software Microsoft Excel, que puede descargarse en el siguiente enlace: <https://zenodo.org/doi/10.5281/zenodo.12535410>

5. Integración cuantitativa y cualitativa del juicio de expertos

El análisis de la concordancia entre jueces se realiza ítem por ítem; para ello, es preciso observar el CVC_{ic} para cada ítem; las puntuaciones se interpretan de acuerdo con lo que se indica en la Tabla 4.

Tabla 4
Interpretación del Coeficiente de Validez de Contenido para cada ítem

Puntuación	Interpretación
$x \geq 0,90$	Validez y concordancia excelente
$0,81 \leq x < 0,90$	Validez y concordancia buena
$0,71 \leq x < 0,80$	Validez y concordancia aceptable
$0,60 \leq x < 0,70$	Validez y concordancia deficiente
$x < 0,60$	Validez y concordancia inaceptable

Fuente: Hernández-Nieto (2011).

Hernández-Nieto (2002) recomienda mantener en el instrumento aquellos ítems con un CVC superior a 0,80. La integración cuantitativa y cualitativa implica que los ítems con una puntuación inferior a 0,80, deberán ajustarse, o bien, eliminarse, según lo indiquen los jueces en la columna “Observaciones” de la *Tabla 2*; además, el equipo investigador evaluará la posibilidad de añadir una nueva dimensión en caso de que los jueces hayan realizado tal sugerencia.

Se recomienda realizar el reporte de resultados en el artículo científico que se realice para difundir las evidencias sobre la construcción y validación del instrumento de medición en cuestión. En el reporte, es deseable que se indique el procedimiento que se siguió para realizar la validez de contenido por juicio de expertos, que puede incluir: a) la definición de perfiles del panel de expertos; b) el número de personas que integraron el panel; c) los resultados del análisis cuantitativo, como el número de ítems que se mantuvieron en el instrumento, los ítems que fueron modificados o eliminados, así como el CVC_{ic} y el CVC_t y d) señalar si se decidió integrar una nueva dimensión, a partir de las sugerencias del panel de expertos.

6. Conclusiones

En el proceso de investigación científica, particularmente en disciplinas como la psicología, las ciencias de la educación y otras ciencias sociales, la presencia de variables latentes plantea la necesidad de crear instrumentos de medición que permitan inferir su existencia. Toda vez que la medición de estos fenómenos plantea cierta complejidad, es fundamental determinar la validez y la confiabilidad de los instrumentos con los que se les evalúa; esto es, asegurarse que el instrumento mide el constructo en cuestión, además de determinar su consistencia y precisión al medir una determinada característica (Fernández-Ballesteros, 2013; Galindo-Domínguez, 2020).

Construir y validar un instrumento de medición, es un proceso riguroso que atraviesa diversas etapas donde los ítems son revisados, corregidos o eliminados, de acuerdo con la evidencia que se obtiene mediante el juicio de expertos, las entrevistas cognitivas, o los indicadores estadísticos derivados de las muestras poblacionales en las que el instrumento se pone a prueba. Lo anterior, con el fin de garantizar mediciones óptimas tanto para el constructo como para la población de interés; esto último, toda vez que las características culturales de ciertos grupos poblacionales, precisa realizar adaptaciones en el instrumento, con el fin de que pueda ser comprendido por la población en la que se lleva a cabo el estudio empírico (García-Saisó *et al.*, 2014; Urpí-Fernández *et al.*, 2020).

Un paso importante en el proceso de construcción de un instrumento de medición consiste en someter los ítems que lo conforman a la evaluación de un panel de especialistas. Este procedimiento permite identificar vacíos conceptuales, ambigüedades semánticas, construcciones gramaticales inadecuadas, posibles sesgos, así como elementos irrelevantes o bien, insuficientes; lo anterior, con el fin de tomar decisiones para efectuar los ajustes

necesarios y garantizar que los ítems, evalúan apropiadamente el constructo de interés (Escobar-Pérez & Cuervo-Martínez, 2008; Galicia Alarcón *et al.*, 2017).

A pesar de que el juicio de expertos comprende un paso importante en el desarrollo de los instrumentos de medición, con frecuencia, las investigaciones no especifican el procedimiento ni los análisis cuantitativos y cualitativos aplicados para la obtención y el procesamiento de las valoraciones de los jueces; de modo que no se brindan detalles sobre aspectos como la selección de expertos, las dimensiones evaluadas y la integración de sus calificaciones y retroalimentaciones para el perfeccionamiento del instrumento, antes de su aplicación.

Por lo tanto, este documento se planteó con el propósito de sintetizar el procedimiento para realizar la validez de contenido por juicio de expertos en la construcción de instrumentos de medición en psicología, mediante una rúbrica que permite emitir una valoración cuantitativa a partir de una escala tipo Likert que va de una puntuación de 1 a 4, así como una valoración cualitativa. Posteriormente, se explicó el procedimiento para llevar a cabo el análisis cuantitativo de las puntuaciones emitidas por el panel de expertos a través del Coeficiente de Validez de Contenido de Hernández-Nieto (2002). Además, se integró una macro de Excel para realizar este cálculo mediante dicho software. Por último, se presentó un planteamiento sobre la integración cuantitativa y cualitativa del juicio emitido por el panel de expertos, junto con la forma de integrar este procedimiento al reporte de resultados. De este modo, la integración de las observaciones de los expertos de manera sistemática mediante una rúbrica con criterios cuali-cuantitativos, agrega transparencia, rigor metodológico y mayor sustento a las modificaciones realizadas al instrumento antes de su aplicación, contribuyendo así a la validez de contenido de la prueba final.

En conclusión, el proceso de construcción y validación de instrumentos para evaluar constructos requiere de múltiples etapas interrelacionadas, desde la conceptualización teórica, pasando por el desarrollo y mejora de los ítems mediante el juicio de expertos, hasta los análisis psicométricos que aportan evidencia cuantitativa sobre su validez y confiabilidad. Estas fases buscan garantizar que el instrumento mida precisa y consistentemente el constructo de interés en una población definida. Por lo tanto, sistematizar el juicio de expertos utilizando rúbricas mixtas, como la propuesta en este documento, proporciona al proceso mayor rigor y trazabilidad. Aunque compleja, la integración cualitativa y cuantitativa de la retroalimentación de los jueces, fortalece la validez de contenido para respaldar que los ítems del instrumento son relevantes y representativos del constructo evaluado; por ello, se recomienda incorporar este procedimiento en el desarrollo de nuevas pruebas psicológicas y realizar el reporte de las decisiones tomadas en la versión final del instrumento.

Agradecimientos

El primer autor, con número de CVU 1040696 agradece al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT), por la beca otorgada para cursar el Doctorado en Psicología con Orientación en Calidad de Vida y Salud en el Centro Universitario del Sur de la Universidad de Guadalajara, en cuyo marco se llevó a cabo este trabajo.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Referencias

- Aiken, L. R. (1980). Content Validity and Reliability of Single Items or Questionnaires. *Educational and Psychological Measurement*, 40(4), 955-959. <https://doi.org/dn3f75>
- Aizpitarte, A., Alonso-Arbiol, I., Van de Vijver, F. J. R., Perdomo, M. C., Galvez-Sobral, J. A., & Garcia-Lopez, E. (2017). Development of a Dating Violence Assessment Tool for Late Adolescence Across Three Countries: The Violence in Adolescents' Dating Relationships Inventory (VADRI). *Journal of Interpersonal Violence*, 32(17), 2626–2646. <https://doi.org/gbscq2>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). *Estándares para Pruebas Educativas y Psicológicas*. American Educational Research Association. <https://doi.org/m5dp>
- Bartlett, L. K., Pirrone, A., Javed, N., & Gobet, F. (2023). Computational Scientific Discovery in Psychology. *Perspectives on Psychological Science*, 18(1), 178-189. <https://doi.org/m5dq>
- Bollen, K. A. (1989). A New Incremental Fit Index for General Structural Equation Models. *Sociological Methods & Research*, 17(3), 303-316. <https://doi.org/cfgdt5>
- Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, 21(2), 230-258. <https://doi.org/dbn>
- Caicedo Cavagnis, E., & Zalazar-Jaime, M. F. (2018). Cognitive interviews: Review, guidelines for use and application in psychological research. *Avaliacao Psicologica*, 17(3), 362-370. <https://doi.org/hz7s>
- Carmines, E. G., & McIver, C. (1981). Analyzing models with unobserved variables: analysis of covariance structures. En G. W. Bohrnstedt, & E. F. Borgatta (Eds.), *Social Measurement: Current Issues* (pp. 65-115). Sage Publications, Inc.
- Cervantes Luna, B. S., Maldonado Rivera, F., Linares Olivas, O. L., & Pesqueira Leal, L. (2017). Introducción a la estadística. En *Estadística aplicada en Psicología y Ciencias de la salud*. El Manual Moderno.
- CIOMS. (2017). *Pautas éticas internacionales para la investigación relacionada con la salud en seres humanos*. <https://tinyurl.com/3prc6vwn>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- Comrey, A. L., & Lee, H. B. (1992). A first course in factor analysis. En *A First Course in Factor Analysis*. Psychology Press. <https://doi.org/c7n8>
- Coolican, H. (2004). *Métodos de investigación y estadística en psicología* (3.ª ed.). El Manual Moderno.
- Danchin, A. (2023). Science, method and critical thinking. *Microbial Biotechnology*, 16(10), 1888–1894. <https://doi.org/m5dr>
- Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). Validez de Contenido y Juicio de Expertos: Una Aproximación a su Utilización. *Avances en Medición*, 6, 27-36.

- Fernández-Ballesteros, R. (2013). Conceptos y metodología básica. En *Evaluación psicológica*. Pirámide.
- Field, A. (2018). *Discovering Statistics using SPSS* (2.ª ed.). SAGE Publications Ltd.
- Fitch, K., Bernstein, S. J., McDonnell, J., & Kahan, J. P. (2001). *The RAND/UCLA Appropriateness Method User's Manual Approved for Public Release* [Informe]. <https://tinyurl.com/5asvvbu2>
- Fleiss, J. L. (1971). Measuring Nominal Scale Agreement Among Many Raters. *Psychological Bulletin*, 76(5), 378-382. <https://doi.org/bzhdfc>
- Fullat, O. (2008). *Filosofías de la Educación* (1.ª ed). Editorial Síntesis.
- 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*, 17(27), 42-53.
- Galindo-Domínguez, H. (2020). *Estadística para no estadísticos: una guía básica sobre la metodología cuantitativa de trabajos académicos* (1.ª ed). <https://doi.org/m5ds>
- García-Saisó, A., Ortega-Andeane, P., & Reyes-Lagunes, I. (2014). Adaptación y Validación Psicométrica de la Escala de Clima Social Organizacional (WES) de Moos en México. *Acta de Investigación Psicológica*, 4(1), 1370-1384.
- Gaxiola Romero, J. C., & Manríquez Betanzos, J. C. (2021). El conocimiento de la historia de la psicología en la evolución de sus constructos teóricos. En J. C. Gaxiola Romero, J. C. Manriquez Betanzos, & L. Y. Yáñez Peñúñuri (Eds.), *Evolución de constructos en psicología. Ejemplos y aportes conceptuales* (pp. 10-28). Qartuppi. <https://doi.org/m5dt>
- Hambleton, R. K. (1984). Validating the test score. En *A Guide to Criterion-Referenced Test Construction* (pp. 199-230). Johns Hopkins University Press.
- Hernández-Nieto, R. (2002). *Contributions to Statistical Analysis*. Universidad de los Andes.
- Hernández-Nieto, R. (2011). *Instrumentos de recolección de datos en ciencias sociales y ciencias biomédicas*. Universidad de los Andes.
- International Test Commission. (2017). *The ITC Guidelines for Translating and Adapting Tests* (2.ª ed.). <https://doi.org/djfsqv>
- Juárez-García, D. M., de Jesús García-Solís, M., & Téllez, A. (2020). Adaptation and Validation of the Health Belief Model Scale for Breast Self-Examination in Mexican Women. *Value in Health Regional Issues*, 23, 30-36. <https://doi.org/gncwqm>
- Kerlinger, F. N., & Lee, H. B. (2002). *Investigación del comportamiento*. McGraw-Hill.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.
- Levine, E. L., Maye, D. M., & Gordon, T. R. (1997). A methodology for developing and validating minimum qualifications (MQs). *Personnel Psychology*, 50, 1009-1023.
- Licon, J. F. M., Oviedo, A. D., Jasso, A. A. S., & Rivera, M. D. (2014). Study of parental models: building an instrument for their exploration. *Revista da Escola de Enfermagem da U S P*, 48(4), 593-600.

- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151–1169. <https://doi.org/gjr929>
- McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, 107(2), 247-255. <https://doi.org/dczz7v>
- Mousazadeh, S., Rakhshan, M., & Mohammadi, F. (2017). Investigation of Content and Face Validity and Reliability of Sociocultural Attitude towards Appearance Questionnaire-3 (SATAQ-3) among Female Adolescents. *Iranian Journal of Psychiatry*, 12(1), 15-20.
- Nunnally, J. C. (1967). *Psychometric theory*. McGraw-Hill.
- Ortega Mohedano, F. (2008). El método Delphi, prospectiva en Ciencias Sociales a través del análisis de un caso práctico. *Revista Escuela de Administración de Negocios*, 64, 31-54. <https://doi.org/gtkqqj>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., & Akl, E. A. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Revista Espanola de Cardiología*, 74(9), 790-799. <https://doi.org/g7qm>
- Park, Y. S., Konge, L., & Artino, A. R. (2020). The Positivism Paradigm of Research. *Academic Medicine*, 95(5), 690-694. <https://doi.org/fm2m>
- Pedrosa, I., Suárez-Álvarez, J., & García-Cueto, E. (2014). Evidencias sobre la Validez de Contenido: Avances Teóricos y Métodos para su Estimación. *Acción Psicológica*, 10(2), 3-20. <https://tinyurl.com/y8ybfdrf>
- Perroca, M. G. (2011). Desarrollo y validación de contenido de la nueva versión de un instrumento para clasificación de pacientes. *Revista Latino-Americana de Enfermagem*, 19(1), 58-66. <https://doi.org/b5tnk6>
- Ramos, C. A. (2015). Los paradigmas de la investigación científica. *Avances En Psicología*, 23(1), 9-17. <https://doi.org/fnv5>
- Rovinelli, R. J., & Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Tijdschrift Voor Onderwijsresearch*, 2(2), 49-60.
- Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94-104. <https://doi.org/fzt54q>
- Ruiz Taborda, J. P., Higuera Gutierrez, L. F., & Cardona Arias, J. A. (2021). Reflexión epistemológica para la investigación de los procesos de determinación social de la salud. *Revista Facultad Nacional de Salud Pública*, 39(1), 1-9. <https://doi.org/m5dx>
- Sireci, S. G., & Geisinger, K. F. (1992). Analyzing Test Content Using Cluster Analysis and Multidimensional Scaling. *Applied Psychological Measurement*, 16(1), 17-31. <https://doi.org/fp5gdw>
- Tourangeau, R. (1984). Cognitive sciences and survey methods. En *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines* (pp. 73-100). National Academy Press.

- Tucker, L. R. (1961). Factor Analysis of Relevance Judgments: An Approach to Content Validity. En A. Anastasi, *Testing Problems in Perspective* (pp. 577-586). American Council on Education.
- Urpí-Fernández, A.-M., Zabaleta-del-Olmo, E., Tomás-Sábado, J., Tambo-Lizalde, E., & Roldán-Merino, J.-F. (2020). Adaptación y validación de un cuestionario para evaluar las prácticas de autocuidado en población infantil sana residente en España. *Atención Primaria*, 52(5), 297–306. <https://doi.org/m5dz>
- Urrutia Egaña, M., Barrios Araya, S., Gutiérrez Núñez, M., & Mayorga Camus, M. (2015). Métodos óptimos para determinar validez de contenido. *Revista Cubana de Educacion Medica Superior*, 28(3), 547-558.
- Vargas Halabí, T., & Mora-Esquivel, R. (2017). Tamaño de la muestra en modelos de ecuaciones estructurales con constructos latentes: Un método práctico. *Revista Electrónica Actualidades Investigativas en Educación*, 17(1), 18-22. <https://doi.org/gdg23z>
- Willis, G. B. (1999). *Cognitive Interviewing. A "how to" guide* [Curso]. Meeting of the American Statistical Association. <https://tinyurl.com/4ukysptb>