

ICrESAI-IMeCI: instrumentos para elegir y evaluar artículos científicos para la investigación y la práctica basada en evidencia³

RESUMEN

Objetivo: el incremento en la producción de conocimiento representa para los científicos una preocupación, dado que diariamente se enfrentan a la decisión de elegir la literatura de mayor mérito científico para respaldar la toma de decisiones para la investigación y la práctica. La presente investigación tiene por objeto construir dos instrumentos registrados⁴, con propiedades psicométricas, validez facial y validez de contenido. **Materiales y métodos:** para el Instrumento de Criterios de Elegibilidad de los Artículos de Investigación para el Metaanálisis (ICrESAI) se proponen los criterios que se deben evaluar para determinar cuáles artículos de investigación deben ser elegibles para una revisión sistemática (RS) o metaanálisis (MA); así mismo, se calculó la fiabilidad interevaluador. El segundo es el Instrumento de Criterios para Evaluar el Mérito Científico de los Artículos Seleccionados para el Metaanálisis (IMeCI), el cual propone criterios que se deben evaluar para calificar el mérito científico y la calidad metodológica de los artículos seleccionados para adelantar una RS. Los cálculos para validez facial y de contenido se hicieron con tres investigadores expertos en metodología de la investigación y en medición. La validez de contenido se calculó con el Índice de Lawshe Modificado (Tristán-López, 2008), y la fiabilidad intercodificador se midió para variables cualitativas con el Índice de Kappa de Cohen. **Resultados:** el instrumento ICrESAI obtuvo un índice de validez de contenido (CVI) de 1,0. Para el instrumento IMeCI, el CVI fue de 0,75. En el instrumento ICrESAI los valores para el índice de fiabilidad o concordancia intercodificador estuvieron entre 0,547 y 0,973, considerados buenos de acuerdo con la interpretación propuesta por Cohen. **Conclusiones:** los dos instrumentos generados cuentan con validez (CVI y Lawshe) para realizar la selección y evaluación del mérito científico de las piezas de investigación. Estas dos herramientas son útiles en el estudio de cualquier área temática y al orientar el uso de la información para la toma de decisiones en la investigación y la práctica basada en evidencia.

PALABRAS CLAVE

Validez de las pruebas, evaluación de la investigación en salud, características de los estudios, enfermería basada en la evidencia, metaanálisis. (Fuente: DeCS-BIREME).

Para citar este artículo / To reference this article / Para citar este artigo

Gómez Ortega, O. R., Amaya Rey, M. C. (2013). ICrESAI-IMeCI: instrumentos para elegir y evaluar artículos científicos para la investigación y la práctica basada en evidencia. *Aquichan*. Vol. 13, No. 3, 407-420.

1 Enfermera. Especialista en Atención de Enfermería en Cuidado Crítico. Universidad Nacional de Colombia, Colombia. orgomez@unal.edu.co

2 Doctora en Filosofía. Profesora Titular, Universidad Nacional de Colombia, Colombia. mcamayad@unal.edu.co

3 Este trabajo se adelantó durante el desarrollo del proyecto de tesis de doctorado titulada Metaanálisis: calidad de vida de personas adultas en pre y post trasplante de órganos. Evidencia para enfermería. Universidad Nacional de Colombia, 2009-2013

4 Registro de Instrumentos ICrESAI e IMeCI: www.derechodeautor.gov.co. Para acceder a los instrumentos escribir a ambas autoras: orgomez@unal.edu.co, mcamayad@unal.edu.co

ICrESAI - IMeCI : Tools for Selecting and Evaluating Scientific Articles for Research and Evidence-based Practice

ABSTRACT

Objective: The increase in the production of scientific knowledge is of concern to scientists, who must decide daily on selecting literature of utmost scientific merit to support decision-making on research and practice. The purpose of this study is to construct two registered instruments⁵ with psychometric properties, face validity and content validity. **Materials and Methods:** The Research Article Eligibility Criteria Instrument (ICrESAI) proposes standards to be assessed to determine which research papers should be eligible for a systematic review (SR) or meta-analysis (MA). Inter-rater reliability was calculated as well. The second instrument involves Criteria for Evaluating the Scientific Merit of Articles Selected for Meta-analysis (IMeCI), which proposes standards to be evaluated to rate the scientific merit and methodological quality of the articles selected for a SR. The estimates for face and content validity were made with three researchers who are experts in research methodology and measurement. Content validity was calculated using the Modified Lawshe Index (Tristán-López, 2008) and inter-coder reliability was measured for qualitative variables using Cohen's Kappa Index. **Results:** The ICrESAI instrument obtained a content validity index (CVI) of 1.0. The CVI for the IMeCI instrument was 0.75. With the ICrESAI instrument, the values for the inter-coder reliability or concordance index were between 0.547 and 0.973, which is considered good according to the interpretation proposed by Cohen. **Conclusions:** Both instruments are valid (CVI and Lawshe) for selecting and evaluating the scientific merit of research articles. These two tools are useful to the study of any subject area and to guide the use of information for decision-making in research and evidence-based practice.

KEY WORDS

Validity of tests, health research evaluation, study characteristics evidence-based nursing, meta-analysis. (Source: DeCS, BIREME).

⁵ ICrESAI and IMeCI registration: www.derechodeautor.gov.co. To access these instruments, contact both authors: orgomez@unal.edu.co, mcamayad@unal.edu.co

ICRESAI-IMECI: instrumentos para escolher e avaliar artigos científicos para a pesquisa e prática baseada em evidência

RESUMO

Objetivos: o aumento na produção de conhecimento representa para os cientistas uma preocupação porque diariamente lidam com a decisão de escolher a literatura de maior mérito científico para respaldar a tomada de decisões para a pesquisa e prática. Esta pesquisa tem por objetivo construir dois instrumentos registrados⁶, com propriedades psicométricas, validade facial e validade de conteúdo. **Materiais e métodos:** para o Instrumento de Critérios de Escolha dos Artigos de Pesquisa para a Meta-análise (ICRESAI), propõem-se os critérios que se devem avaliar para determinar quais artigos de pesquisa devem ser eleitos para uma revisão sistemática (RS) ou meta-análise (MA); além disso, calculou-se a fiabilidade interavaliador. O segundo é o Instrumento de Critérios para Avaliar o Mérito Científico dos Artigos Seleccionados para a Meta-análise (IMECI), o qual propõe critérios que se devem avaliar para qualificar o mérito científico e a qualidade metodológica dos artigos seleccionados para realizar uma RS. Os cálculos para validade facial e de conteúdo foram feitos com três pesquisadores especialistas em metodologia da pesquisa e em medição. A validade de conteúdo foi calculada com o Índice de Lawshe Modificado (Tristán-López, 2008), e a fiabilidade intercodificador foi medida para variáveis qualitativas com o Índice de Kappa de Cohen. **Resultados:** o instrumento ICRESAI obteve um índice de validade de conteúdo (CVI) de 1,0. Para o instrumento IMECI, o CVI foi de 0,75. No instrumento ICRESAI os valores para o índice de fiabilidade ou concordância intercodificador estiveram entre 0,547 e 0,973, considerados bons de acordo com a interpretação proposta por Cohen. **Conclusões:** os dois instrumentos gerados possuem validade (CVI e Lawshe) para realizar a seleção e avaliação do mérito científico das peças da pesquisa. Estas duas ferramentas são úteis no estudo de qualquer área temática e na orientação do uso da informação para a tomada de decisões em pesquisa e prática baseada em evidência.

PALAVRAS-CHAVE

Validade dos testes, avaliação da pesquisa em saúde, características dos estudos, enfermagem baseada em evidências, metanálise (Fonte: DeCS, BIREME).

⁶ Registro de Instrumentos ICrESAI e IMECI: www.derechodeautor.gov.co. Para acessar os instrumentos, escreva a ambas as autoras: orgomez@unal.edu.co, mcamayad@unal.edu.co.

Introducción

En las últimas décadas se ha evidenciado un incremento importante en el número de publicaciones científicas generadas desde diferentes disciplinas y áreas temáticas. Como consecuencia, el interés en ejes temáticos específicos ha impulsado la formación de sociedades científicas y académicas, que día a día trabajan para ofrecer a la comunidad un conocimiento nuevo e innovador que permita entender mejor los problemas o fenómenos que se presentan en la práctica profesional (1, 2).

Este incremento en el volumen de difusión de la producción de conocimiento ha despertado en los científicos y profesionales un interés primordial en poder contar con herramientas para elegir la literatura más relevante, que oriente la toma de decisiones y resolución de problemas relacionados con el ejercicio profesional de diferentes disciplinas. Por ello, el objetivo de esta investigación es la creación y validación de dos instrumentos que respondan a la necesidad de disponer de herramientas útiles para facilitar la realización de revisiones sistemáticas (RS) o investigaciones secundarias como metaanálisis. Las revisiones sistemáticas tienen como propósito consolidar de forma objetiva y condensada el estado del arte de un área temática, basándose en la crítica y las evaluaciones de los resultados de investigaciones u otros documentos publicados previamente a las RS o metaanálisis. Una revisión sistemática de una área temática exige que el investigador, a partir de la búsqueda del planteamiento de una pregunta de investigación, realice una selección amplia de la literatura, de varias fuentes de información, bases de datos, a fin de consolidar la información disponible para luego presentarla de manera clara, precisa y comprensible, a través de un artículo científico (3).

Ahora bien, la sistematización y metodología con que se realiza la selección y crítica de las publicaciones o los estudios de investigación, el análisis estadístico, metodológico y conceptual de los contenidos, y la forma en que se consolidan los resultados de las investigaciones primarias, hacen posible la realización de diferentes tipos de revisiones sistemáticas de la literatura que se pueden clasificar de tres formas (3- 5):

- a. La revisión narrativa, temática o documental: también se conoce como revisión general. En ella se organizan ejes temáticos del conocimiento publicado, con el propósito de dar una visión general del problema en investigación. Este tipo de revisión motiva discusiones, presenta nuevas ideas de investigación y permite aclarar significados de algunos conceptos.
 - b. La revisión sistemática cualitativa, metasíntesis o revisiones integrativas (RI): analiza estudios con abordaje cualitativo, sigue un proceso sistemático y metódico, permite entender los fenómenos y los procesos, además de esclarecer significados de los conceptos.
 - c. Revisión sistemática cuantitativa: también conocida como metaanálisis (MA), es la revisión de literatura que incluye solo estudios de abordaje cuantitativo, que permiten dar respuesta a una sola pregunta de investigación a partir de la consolidación de hallazgos obtenidos en investigaciones primarias. Los datos centrales de análisis son los resultados estadísticos de estos estudios primarios.
- Sigue un método sistemático que permite identificar, escoger y criticar la mejor evidencia existente sobre un problema específico. Los resultados obtenidos de este tipo de revisión conforman el nivel más alto de evidencia.
- Las revisiones sistemáticas corresponden a un tipo de investigación metodológica o secundaria, que hoy ha ido ganando espacio en el mundo de la ciencia. En la actualidad se destacan grupos de profesionales que se han dedicado a orientar sobre la metodología correcta para llevar a cabo este tipo de estudios, por ejemplo: la Colaboración Cochrane, el instituto Joanna Briggs y el grupo JAMA de Medicina Basada en Evidencia (5- 7). Estos grupos de investigadores defienden un nuevo paradigma dentro de la investigación que se conoce en la actualidad con el concepto de práctica basada en la evidencia, en el que se busca demostrar que a partir del conocimiento construido es posible encontrar los mejores argumentos científicos para lograr la comprensión y resolución de los problemas de la investigación y de la práctica.
- Como cualquier otro tipo de investigación, la realización de una revisión sistemática cuantitativa, o metaanálisis (MA), debe acoger todas las etapas propias del desarrollo de un proyecto de investigación: el planteamiento de una pregunta o problema de investigación, la definición de unos objetivos e hipótesis, la búsqueda de la literatura que permita tener una visión más amplia y precisa del fenómeno que se está investigando. Una vez se ha dado cumplimiento a cada una de las anteriores etapas, el proceso que se debe seguir es la selección de la literatura más relevante sobre el tema estudiado y la cualificación o evaluación de mérito científico de la literatura seleccionada, para luego lograr la consolidación del conocimiento y decidir, a partir de los hallazgos

obtenidos, cuáles son las posibilidades de solución del problema y tomar decisiones frente a la práctica (6, 7).

Un aspecto importante sobre el que se sigue adelantando investigación, y no hay unificación en los grupos de científicos, es cuál es el mejor método para evaluar el conocimiento existente sobre una temática. La evaluación del conocimiento construido hace parte de la práctica habitual de los profesionales y debe ser considerada un aspecto primordial al realizar una revisión sistemática de la literatura, ya que cualificar la calidad científica de los estudios existentes permite tener un panorama general de lo que se conoce o desconoce de un fenómeno específico.

En la actualidad existen organismos y autores que proponen estrategias para realizar la lectura y evaluación crítica de la literatura (*critical appraisal*), entre los que sobresalen: Agency for Health Care Research and Quality, Cochrane, Critical Appraisal Skills Programme Español (CASPe) (8), Normas para la Crítica de Artículos Originales en Salud (9, 10), bases para la revisión crítica de artículos médicos (11, 12), el sistema GRADE para evaluar la calidad de la evidencia y la graduación de la fuerza de las recomendaciones (13).

Por otra parte, Gómez-Gómez *et al.*, Des Jarlais *et al.*, Berra *et al.*, y CASPe proponen que la evaluación crítica de los estudios debe dividirse en las siguientes secciones: 1) evaluación de la estructura básica de un artículo, 2) evaluación de aspectos específicos de acuerdo con los principales diseños de investigación, 3) evaluación del contenido del estudio, 4) valoración de las relaciones de causalidad reportadas y 5) evaluar si la selección de las pruebas estadísticas es adecuada (8, 11-13).

Ante esta diversidad de propuestas para determinar la calidad o el mérito científico del conocimiento construido sobre un área temática específica surge la inquietud acerca de cuál es el mejor método para evaluar si un estudio constituye un insumo apropiado para un metaanálisis (MA), retomando que el MA se define como un estudio que utiliza métodos cuantitativos para analizar, combinar e integrar de manera comprensiva los resultados obtenidos en investigaciones científicas realizadas sobre un problema concreto (14-16). Al igual que cualquier otro tipo de investigación, sigue un proceso sistemático, metodológico y riguroso, que garantiza validez, precisión, objetividad y reproducibilidad de los resultados y conclusiones (15). Sus aplicaciones incluyen la evaluación de la eficiencia de terapias y la planificación de nuevos estudios, entre otros (14, 17).

Dentro de las fases descritas en la realización de un metaanálisis, la selección de los estudios y cualificación de estos es un paso fundamental. Al no contar con instrumentos con propiedades psicométricas para determinar cuáles estudios incluir en un MA y cómo evaluar de manera objetiva el mérito científico, el presente estudio tiene por objeto la construcción de dos instrumentos: ICrESAI, que facilita determinar cuáles artículos científicos son aptos para conformar el metaanálisis, así como IMeCI, que busca determinar cuál es el mérito científico y la calidad metodológica de las publicaciones científicas.

Este proceso demandó realizar una revisión de la literatura sobre temas que se consideran primordiales a fin de responder a las siguientes preguntas: 1) ¿cómo decidir qué artículos elegir para ser incluidos en un MA o una revisión sistemática? y 2) ¿qué criterios permiten establecer el mérito científico de los artículos que resultan elegidos? Para dar respuesta a estos dos interrogantes se consideraron a) el análisis de la estructura que debe tener el informe escrito de una investigación, y b) el cumplimiento del rigor conceptual, metodológico y empírico de la investigación seleccionada.

Marco conceptual

A continuación se presentan los conceptos centrales: área temática, crítica de la literatura, criterios de calidad, mérito científico, validez facial, validez de contenido, fiabilidad intercodificador, investigación metodológica-instrumental y pregunta o ítem:

Área temática: consolidado de conocimientos afines, unificados conceptual y teóricamente, existentes sobre un campo de trabajo o problema por investigar en una disciplina (18). Se usa con múltiples fines en la investigación y la práctica basada en la evidencia.

Crítica de la literatura: conlleva el análisis y la valoración de los escritos científicos. Permite determinar la coherencia entre el planteamiento problemático, los objetivos, la metodología y los resultados presentados en una investigación, así como los artículos que por la solvencia científica contribuyen a la toma de decisiones para la investigación y la práctica (9, 10, 17).

Criterios de calidad o criterios para evaluar el mérito científico: corresponden a preguntas, pautas o ítems de evaluación que garantizan alta rigurosidad metodológica, la validez interna y la utilidad de los resultados de una investigación en el

abordaje y la comprensión de un problema existente en una disciplina (3, 7, 9, 10, 11).

Mérito científico: resultado afirmativo del cumplimiento de la crítica rigurosa desde lo conceptual, metodológico y empírico, dependiendo del fenómeno en estudio. Es útil para la sistematización y el ordenamiento de los antecedentes de investigación y de lo que se conoce del área temática como referente para las innovaciones de la práctica (18).

Validez facial: también llamada validez de apariencia o validez de expertos. Se refiere al grado en que aparentemente un instrumento mide la variable en cuestión, de acuerdo con “voces calificadas”. Según Sánchez, Echeverry y Gómez, la validez facial es el grado en que una situación de medida parece ser válida desde el punto de vista del observador. “No supone un concepto estadístico, sino que depende del acuerdo que el juicio de los expertos haga sobre la pertinencia de los ítems de la escala” (19, 20).

Según Echeverri (20), en validez facial cada experto debe evaluar para cada ítem:

- Claridad: “Frase o ítem con lenguaje y redacción fácil de comprender”.
- Precisión: “Frase o ítem puntual y exacto que no da lugar a dudas”.
- Comprensión: “Entendimiento de lo que se quiere decir al leer el ítem”.

Validez de contenido: grado en que un instrumento refleja un dominio específico de contenido de lo que mide. Es el grado en el que la medición representa al concepto o variable medida (21-24). Cuando se desarrolla o se quiere evaluar una escala, se debe pretender que los ítems o las preguntas de la escala cubran adecuadamente todos los dominios (o las principales categorías o dimensiones) de la entidad (concepto) que se quiere medir. Un dominio, dimensión o categoría es un grupo de características que se encuentran comúnmente presentes en la entidad; los ítems o preguntas son unidades de exploración que permiten evaluar la presencia de estos atributos en los diferentes dominios, dimensiones o categorías (23).

Existen múltiples pruebas estadísticas con las cuales se puede llevar a efecto el análisis de la validez de contenido (22). Para el presente estudio el análisis estadístico de validez de contenido se realizó con la prueba de Lawshe modificado de Tristán-López

(23, 24), que califica cada pregunta o ítem como esencial, útil o no necesario, y describe que solo las preguntas o ítems evaluados como *esenciales* son relevantes para el concepto o dimensión, atributo(s) que se mide(n).

Evaluación fiabilidad intercodificador: corresponde a la aplicación de pruebas estadísticas que tienen por objeto maximizar la fiabilidad, controlar la posibilidad de errores y asegurar que los datos por analizar son correctos, de acuerdo con lo que previamente se ha planeado evaluar y extraer de cada estudio. Es aplicada por un mínimo dos jueces/evaluadores o codificadores y verificada, según se necesite, por un tercer juez/evaluador o codificador. Este cálculo de la fiabilidad en la codificación de las características de los estudios es a su vez una medida de la calidad de una revisión sistemática o metaanálisis (25-28).

Investigación metodológica-instrumental: la investigación metodológica-instrumental se refiere a investigación controlada sobre las formas para obtener, organizar y analizar datos. Los estudios metodológicos abordan la creación, el desarrollo, la prueba, la validación, la evaluación y la estandarización de los instrumentos o técnicas de investigación. El objetivo es desarrollar un instrumento confiable, eficaz y funcional, que otros investigadores o profesionales puedan utilizar para evaluar el (los) concepto(s), y así lograr sus propósitos para la investigación y la práctica (29, 30).

Pregunta o ítem: cada una de las partes o unidades de que se compone una prueba, un test, un cuestionario (31). El término ítem se utiliza en evaluación para referirse a una pregunta en una *prueba de corrección objetiva*. En muchas ocasiones, en español el término ítem puede utilizarse indistintamente como sinónimo de pregunta. Se pueden clasificar los ítems con referencia al tipo de respuesta que deberá dar la persona encargada de responder. Así, tenemos *ítems de respuesta abierta* e *ítems de respuesta cerrada*. En los instrumentos propuestos los ítems son de respuesta cerrada; la persona que evalúa debe marcar una alternativa o seleccionar una opción entre las dos o más que se proponen (32).

Materiales y métodos

Para los dos instrumentos: ICrESAI e IMeCI

El estudio corresponde a una investigación metodológica-instrumental, que tuvo por objetivos, en primer lugar, diseñar

dos instrumentos: Instrumento de Criterios de Elegibilidad para la Selección de Artículos de Investigación (ICrESAI), que permite determinar cuáles artículos son elegibles para ser parte de una revisión sistemática, e Instrumento de Criterios para Evaluar el Mérito Científico de los Artículos de Investigación Seleccionados para el Metaanálisis (IMeCI), en el cual se evalúa la calidad de las piezas de investigación sobre un área temática específica. Posteriormente se presenta la evaluación de la validez facial y la validez de contenido de los dos instrumentos, así como los índices de fiabilidad en la codificación o evaluación interevaluador.

Para el diseño y la validez de los dos instrumentos propuestos de criterios de elegibilidad y evaluación del mérito científico de los artículos para un metaanálisis se utilizó la metodología propuesta por LoBiondo-Wood y Haber (33), la cual contempla:

- La definición del constructo que se va a medir.
- La formulación de los ítems del instrumento.
- El desarrollo de instrucciones para los expertos.
- La validez del instrumento.
- Este estudio se realizó en dos fases:
 - Fase A: construcción de los instrumentos.
 - Fase B: validez facial o de apariencia y validez de contenido del instrumento.

Construcción de los dos instrumentos

Previo a la creación de los instrumentos ICrESAI e IMeCI se llevó a cabo una extensa revisión de la literatura para evaluar los parámetros propuestos por autores y grupos de científicos organizados, sobre aspectos centrales por evaluar en el momento de realizar la selección y la lectura crítica para definir el mérito científico de una publicación. Se encontraron diversas propuestas (5, 6, 9, 10, 13, 14), pero ninguna de ellas ofrecía un instrumento con propiedades psicométricas evaluadas y pautas de interpretación de los resultados. Correspondían más a una lista de chequeo donde se presentan los contenidos y la calidad de los contenidos que deben estar presentes en un artículo, que permitan dar una evaluación y calificación objetiva de los productos de investigación científica publicados. Por ello, se procedió a determinar las preguntas o pautas y puntajes asignados que permitan obtener la calificación definitiva.

Para ambos instrumentos, los ítems definitivos se crearon a partir de la revisión de diferentes listas de chequeo generadas para evaluar, con base en la lectura crítica, la calidad de los estudios científicos. Luego estos fueron organizados en categorías que correspondieron con los apartados o contenidos que hacen parte de la estructura de una publicación científica, que a la vez se corresponden con las fases de desarrollo de un proyecto de investigación (9, 10). Se generó una escala de calificación para cada ítem. Para los instrumentos propuestos correspondió a las respuestas SÍ (si en el artículo se cumple con lo evaluado, se le asignó un valor de uno 1), NO (cuando en el artículo no se cumple con lo evaluado, se asignó un valor de cero 0) o NO APLICA (NA) (corresponde a una opción que solo es susceptible de responder en aquellos artículos con características específicas en el diseño, como estudios comparativos, correlacionales o experimentales, etc.). Para la opción NA no se asignó un puntaje de calificación (no suma en el balance total).

Una vez se tuvieron las primeras versiones de los instrumentos, se remitieron a un experto en lingüística, quien procedió a revisar la semántica y sintaxis de cada pregunta o ítem. Posteriormente, los dos instrumentos se llevaron a un juicio de expertos, quienes tenían por función en primera instancia evaluar la validez facial de cada ítem o pregunta. Siguiendo las recomendaciones de Sánchez y Echeverri (20, 34), a cada pregunta o ítem se le evaluó la claridad, precisión y comprensión.

El grupo de expertos evaluadores estuvo conformado por tres profesionales calificados con formación académica de maestría o doctorado, con experiencia académica y profesional, autores de artículos científicos, con experiencia en metodología de investigación y medición de fenómenos en salud, en crítica de publicaciones científicas, que actúan como pares evaluadores en comités editoriales de revistas científicas indexadas. Uno de ellos, además de las anteriores competencias, tenía conocimiento en análisis estadísticos de estudios metaanalíticos.

Análisis estadístico

Para el cálculo de la validez de contenido, la prueba estadística utilizada fue el Lawshe modificado de Tristán-López (24). Este autor propone una modificación al modelo de Lawshe original, argumentando que una de las mayores dificultades para evaluar la validez de contenido de una escala es poder reunir el número suficiente de expertos; con su propuesta se demuestra que en

mediciones con un mínimo de participación de expertos es posible obtener valores válidos en cuanto al contenido; señala que el valor de referencia mínimo aceptable para la razón de validez de contenido (CVR) es de 0,5823, el cual será estable independientemente del número de expertos que participen en el proceso. Por otra parte, el índice de validez de contenido de la prueba (CVI) se calcula tomando todas las CVR que dieron valores iguales o superiores a 0,5823, sumándolos y luego dividiéndolos entre el número de valores aceptados.

La CVR busca determinar, en cada una de las categorías, cuáles son los ítem más relevantes de medir para garantizar que un constructo o concepto se calcula adecuadamente. Desde la propuesta de Tristán cada ítem puede ser evaluado como esencial (el ítem es importante e indispensable), útil (importante aunque no indispensable) o no necesario (ni importante ni indispensable). Solo se consideraron los ítems que recibieron calificación de esencial por el mayor número de expertos y mostraron una CVR mayor de 0,5823. Luego todos los ítems que superaron este valor se sumaron y promediaron. El valor obtenido correspondió al CVI (índice de validez de contenido de cada una de las dimensiones y de todo el instrumento). Las fórmulas para hallar estos dos valores se explican a continuación:

CVR: *Número de acuerdo (# evaluadores que califican el ítem como esencial).*

Número total de evaluaciones realizadas (para nuestro ejemplo, 3 evaluadores).

Valores por encima de 0,5823 son considerados aceptables y el ítem se retiene como parte del instrumento.

CVI: *Sumatoria de todos los valores de los ítems con CVR \geq 0,5823.*

Número total de ítems con evaluación aceptable (CVR \geq 0,5823).

La fórmula para calcular la fiabilidad o acuerdo entre dos evaluadores depende de la escala en que es medida cada una de las variables o características codificadas para cada estudio (25-28). Para características *nominales, categóricas o cualitativas* se utiliza como indicador de la fiabilidad el coeficiente Kappa de Cohen (caracterizado con el símbolo k). Este coeficiente permite

cuantificar el nivel del acuerdo interobservador para disminuir la subjetividad del método utilizado y descartar que el grado de acuerdo se deba al azar. Se define como un estadístico de concordancia entre dos investigadores que corrige el azar. Al ser una probabilidad, toma su valor en el intervalo (0, 1). La máxima concordancia posible corresponde a $\kappa = 1$. El valor $\kappa = 0$ se obtiene cuando la concordancia observada es precisamente la que se espera a causa exclusivamente del azar. Si la concordancia es mayor que la esperada simplemente a causa del azar, $k > 0$, mientras que si es menor, $k < 0$. El mínimo valor de k depende de las distribuciones marginales.

Kappa es “un corrector de la medida de acuerdo interevaluador” que permite verificar que el nivel de acuerdos exceda los niveles de suerte.

Se representa según la fórmula:

$$k = \frac{Po - Pa}{1 - Pa}$$

Donde:

Po : proporción de concordancia observada realmente y se calcula sumando las marcas que representan la concordancia y dividiendo entre el número total de ellas.

Pa : proporción esperada por azar, se calcula sumando las probabilidades de acuerdo por azar para cada categoría.

Donde los valores k obtenidos deben interpretarse de la siguiente manera:

- < 0,20: pobre
- 0,21 y 0,40: débil
- 0,41 y 0,60: moderada
- 0,61 y 0,80: buena
- 0,81 y 1,00: muy buena

Un cálculo que reporte un coeficiente Kappa de 0,4 puede considerarse como el límite de fiabilidad aceptable de una prueba.

Resultados

Los resultados del estudio se presentan por separado para cada uno de los dos instrumentos diseñados y validados.

Construcción y validez del Instrumento de Criterios de Elegibilidad de los Artículos de Investigación para el Metaanálisis (ICrESAI)

Una vez concluida la revisión de literatura, el instrumento ICrESAI de criterios de elegibilidad quedó compuesto por 5 categorías que se mantienen hasta la versión final del instrumento, y 11 ítems o preguntas que se fueron definiendo a partir de las recomendaciones y la evaluación de los expertos. Para cada ítem se revisó la semántica y sintaxis por un experto en lingüística y se realizaron las correcciones sugeridas. Posteriormente, los tres profesionales expertos en medición evaluaron la validez facial de cada una de las dimensiones e ítems. Se calculó el índice de acuerdo, se realizaron los ajustes y las recomendaciones dadas, y se les devolvió a los expertos con las correcciones realizadas hasta tener una última versión clara, precisa y comprensible de los tres evaluadores.

Una vez concluida la validez facial del instrumento se procedió a medir la validez de contenido; para ello se aplicó la prueba estadística de Lawshe modificada de Tristán-López, 2008, la cual arrojó la conveniencia de retirar un ítem de la dimensión muestra por considerarse *no esencial*. El instrumento finalmente quedó constituido por cinco categorías y diez ítems, de los cuales los nueve primeros eran de obligatorio cumplimiento para que el artículo en evaluación se considerara elegido para la inclusión en el metaanálisis. El segundo ítem de la dimensión de resultados solo era evaluable en estudios experimentales, correlacionales o comparativos.

En consulta realizada a un profesional experto en medición y en la metodología de metaanálisis, se recomendó para ICrESAI modificar la dimensión Instrumento de medición, retirando el ítem de propiedades psicométricas con el fin de evitar el sesgo de selección que podría conllevar la exclusión de gran parte de los artículos disponibles por fijar criterios de elección muy rígidos. Se recomienda trasladar este ítem fundamental al instrumento de evaluación del mérito científico.

Por lo anterior, la versión definitiva del instrumento se constituyó de cinco dimensiones y nueve ítems (Tabla 1). Los primeros siete ítems se consideraron de obligatorio cumplimiento (entiéndase por obligatorio cumplimiento aquellos ítems que deben estar presentes en las piezas de investigación para ser parte del

metaanálisis o de la revisión sistemática). Para los ítems 8 y 9 se consideró que el cumplimiento de uno de ellos, en conjunto con los anteriores siete, era suficiente para tomar la decisión de incluir o excluir un estudio del metaanálisis.

En el cálculo de la validez de contenido se tuvo en cuenta que todas las dimensiones e ítems propuestos eran esenciales para decidir qué literatura elegir como parte de una revisión sistemática. Se obtuvo un índice de validez de contenido (CVI) del ICrESAI de 1,0.

Tabla 1. Resultados de la validez de contenido de los instrumentos de criterios de elegibilidad de los estudios (ICrESAI) (5 categorías y 9 ítems)

	Categoría evaluada	Índice Lawshe por categoría
1	Objetivos del estudios (1 ítem)	1
2	Muestra (2 ítems)	1
3	Instrumento de medición (2 ítems)	1
4	Procedimiento de recolección y análisis de los datos (2 ítems)	1
5	Presentación de los resultados (2 ítems)	1
	Índice de validez del instrumento ICrESAI contenido (CVI)	1

De los resultados se concluye que el instrumento diseñado cuenta con validez aparente y de contenido adecuada, y puede ser utilizado para decidir cuáles piezas de la literatura cumplen con los requisitos para ser parte de la investigación metodológica: metaanálisis.

Para el cálculo del índice de fiabilidad o concordancia interevaluador del instrumento ICrESAI, compuesto por ítems medidos en una escala dicotoma o nominal, la prueba estadística aplicada fue el Kappa de Cohen. La medición se realizó en 263 estudios considerados elegibles para un metaanálisis que abordaba el estudio de la calidad de vida de las personas adultas que experimentan un trasplante de órganos. Los resultados de la prueba de fiabilidad fueron (Tabla 2):

El Kappa de Cohen es una prueba que presenta resultados de difícil interpretación para los ítems en que las respuestas se

Tabla 2. Fiabilidad interevaluador del instrumento ICrESAI
Artículos evaluados: 263

Ítem No.	Categoría / Ítem	Índice Kappa Cohen	P	# Acuerdo	% Acuerdo
1	Objetivos del estudio: Ítem - Elegibilidad 1	0,973	0,000*	255	96,96
2	Muestra: Ítem - Elegibilidad 2	-0,004	0,951	261	99,24
3	Muestra: Ítem - Elegibilidad 3	0,853	0,000*	261	99,24
4	Instrumento de Medición: Ítem - Elegibilidad 4	0,665	0,000*	262	99,62
5	Instrumento de Medición: Ítem - Elegibilidad 5	-0,005	0,930	260	98,86
6	Procedimiento de recolección y análisis de los datos: Ítem - Elegibilidad 6	0,000		262	99,62
7	Procedimiento de recolección y análisis de los datos: Ítem - Elegibilidad 7	0,855	0,000*	262	99,62
8	Presentación de los resultados: Ítem - Elegibilidad 8	0,547	0,000*	251	95,44
9	Presentación de los resultados: Ítem - Elegibilidad 9	0,785	0,000*	237	90,11
10	Elegibilidad definitiva	0,967	0,000*	259	98,48

concentran en una única opción, o donde una de las marginales tiene valor de cero. Da resultados bajos o negativos, aunque el nivel de porcentaje de acuerdo sea alto. Es por ello que se presenta, además del índice de Kappa, el porcentaje de acuerdo entre las respuestas emitidas por los dos evaluadores. En general se obtuvieron índices de Kappa de Cohen buenos y muy buenos para todos los ítems evaluados.

Construcción y validez del Instrumento de Criterios para Evaluar el Mérito Científico de los Artículos Seleccionados para el Metaanálisis (IMeCI)

En el diseño del presente instrumento, inicialmente a partir de una robusta revisión de la literatura (5-14, 17), se definieron las

categorías e ítems considerados importantes para la evaluación crítica de literatura científica. El instrumento de evaluación del mérito científico inicialmente estaba constituido de la siguiente manera: 10 categorías, que a su vez correspondían a los apartados de la estructura de un artículo científico, y 78 ítems, con los cuales se evaluaba la solvencia conceptual, metodológica y empírica de la investigación profundizada en un artículo científico.

Dimensiones e ítems del instrumento IMeCI:

1. Pautas generales a evaluar (6 ítems)
2. Pautas para evaluar título (3 ítems)
3. Pautas para evaluar autores (4 ítems)
4. Pautas para evaluar resumen (6 ítems)
5. Pautas para evaluar introducción y marco teórico (24)
6. Pautas para evaluar diseño (5 ítems)
7. Procedimiento de recolección de la información (1 ítem)
8. Pautas para evaluación de los resultados (14 ítems)
9. Pautas para evaluar discusión, conclusiones y recomendaciones (9 ítems)
10. Pautas para evaluar bibliografía (6 ítems)

El instrumento IMeCI se sometió a la revisión semántica y sintáctica de un experto en lingüística y, una vez realizados los cambios sugeridos, se sometió al juicio de expertos para evaluar la validez facial de acuerdo con los parámetros dados por Echeverri (20). El instrumento fue reconstruido siguiendo las recomendaciones para validez aparente: se retiraron los ítems considerados repetitivos o redundantes y se especificaron las dimensiones en las cuales iba a ubicarse cada ítem. Después de la prueba de validez facial, el primero quedó constituido por 11 categorías o dimensiones y 61 ítems. Se organizó la versión preliminar del instrumento y se procedió a calcular la validez de contenido de acuerdo con la propuesta de Tristán-López, 2008, obteniéndose para el IMeCI un CVI aceptable de 0,778.

La versión final de los instrumentos quedó conformada de la siguiente manera: 11 categorías o dimensiones y 55 preguntas. Pero dada la recomendación de los expertos metodológicos en metaanálisis, para dar mayor validez a los resultados obtenidos tras la aplicación del instrumento se decidió incluir un nuevo ítem

para la categoría de pautas a fin de evaluar la muestra que varía el porcentaje de atrición (pérdida) de la muestra en el estudio. Además se generó una nueva categoría con un ítem que correspondió a evaluación psicométrica del (los) instrumento(s) de medición utilizado(s) en el estudio, que inicialmente hacía parte del instrumento de criterios de elegibilidad (ICrESAI). También se suprimió un ítem de la dimensión de muestra por considerarse duplicado. La versión definitiva del instrumento quedó constituida por 12 dimensiones o categorías y 56 preguntas o ítems.

Para la última versión del instrumento IMeCI se recalculó el índice de validez de contenido y se obtuvo un CVI aceptable de 0,75.

Tabla 3. Validez de contenido del instrumento para evaluar el mérito científico de los estudios para el metaanálisis (IMeCI) (12 categorías y 56 ítems)

Categoría evaluada	Índice Lawshe por categoría
Pautas generales claves a evaluar (4 ítems)	0,74
Pautas para evaluar el título (2 ítems)	0,67
Pautas para evaluar autores (2 ítems)	0,67
Pautas para evaluar el resumen (6 ítems)	0,84
Evaluación de la introducción y marco teórico (11 ítems)	0,84
Pautas para evaluar el diseño (4 ítems)	0,74
Pautas para evaluar instrumento de medición (1 ítem)	0,67
Pautas para evaluar muestra (7 ítems)*	0,73
Pautas para evaluar el procedimiento de recolección y análisis de la información (3 ítems)	1
Pautas para evaluación de los resultados (8 ítems)	0,79
Pautas para evaluar discusión, conclusiones y recomendaciones (6 ítems)	0,67
Pautas para evaluar las referencias (2 ítems)	0,67
Índice total de validez de contenido	0,75
* Este cálculo del Índice de Lawshe de esta categoría se realizó solo con 6 ítems, dado que el ítem que evalúa el porcentaje de pérdida de tamaño de muestra (atrición de la muestra) se incluyó después, durante la fase de aplicación del instrumento de los estudios seleccionados para un metaanálisis.	

La escala de calificación para cada ítem propuesto en el instrumento correspondió a las siguientes opciones de respuesta: SÍ (si en el artículo se cumple con lo evaluado, se le asignó un valor de uno 1), NO (cuando en el artículo no se cumple con lo

evaluado, se asignó un valor de cero 0) o NO APLICA (NA) (corresponde a una opción que solo es susceptible de responder en los artículos con características específicas en el diseño, como estudios comparativos, correlacionales o experimentales, etc.). Para la opción NA, no se asignó un puntaje de calificación (no suma en el balance total).

Se consideraron como ítems especiales que puntúan, según su alcance, aspectos como tipo de publicación de 1 a 4, tipo de muestra de 1 a 2, diseño de 1 a 2 y propiedades psicométricas del instrumento de medición de 1 a 3. A estos ítems se les otorgó un valor diferente de puntuación (0, 1 o no aplica), por ser los que garantizan el valor metodológico. Los valores asignados a cada opción de respuesta se explican en el manual de diligenciamiento del instrumento. La puntuación definitiva del artículo evaluado se obtiene de la suma de las respuestas obtenidas en cada ítem examinado. Una vez evaluados todos los ítems, se suman los valores obtenidos por el total de ítems que aplican. El máximo valor de calificación está dado por la suma de los ítems estimados como positivos más los negativos (todos los ítems que aplican a cada artículo). Así, se evaluó el mérito científico de acuerdo con los percentiles 25-50 y 75. Se definieron para ello tres categorías para calificar la calidad del mérito científico de la pieza de investigación: alta, media o baja, de acuerdo con la proporción de la sumatoria de los ítems que cargaron puntos, sobre el total de aquellos que aplican. Si la proporción está por debajo del 50% de los que aplican, es bajo; si está entre 50-75% de los que aplican, es medio, y si es superior al 75% de los que aplican, es alto el mérito científico.

El IMeCI cumple con propiedades psicométricas y puede ser utilizado para realizar la crítica del mérito científico de artículos de investigación seleccionados sobre un área temática específica, así como para valorar la calidad del conocimiento basado en la rigurosidad conceptual, metodológica y empírica con que una investigación es realizada.

El instrumento generado propone dos formas de interpretar los resultados. El puntaje obtenido de la evaluación de la totalidad de los ítems será un indicador del mérito científico del estudio y 18 ítems del instrumento se integraron para obtener una medida de la calidad metodológica del estudio. Estos ítems se distribuyen dentro de cada categoría de la siguiente manera (Tabla 4):

Tabla 4. Categorías e ítems para evaluar la calidad metodológica de los estudios

Categoría evaluada	No. de ítems evaluados
Evaluación de la introducción y marco teórico	2
Pautas para evaluar el diseño	3
Pautas para evaluar el instrumento de medición	1
Pautas para evaluar muestra	5
Pautas para evaluar el procedimiento de recolección y análisis de la información	2
Pautas para evaluación de los resultados	3
Pautas para evaluar discusión, conclusiones y recomendaciones	2
Total de ítems que definen la calidad metodológica	18

Conclusiones y proyección futura

La aplicación de los instrumentos ICRESAI e IMeCI permitirá determinar la selección, el mérito científico y la calidad metodológica de los artículos originales existentes sobre un área temática especificada según el fenómeno o fenómenos de interés, objeto de análisis en una revisión sistemática. Pero se considera que puede ser útil para leer, criticar o analizar de forma más profunda artículos originales de abordaje cuantitativo y diferente alcance investigativo.

Referencias

1. Aguillo IF, Correa E, Martínez C. Producción científica y tecnológica nacional y presencia web de instituciones académicas en Iberoamérica. *Interciencia* 2010; 35(2):92-98.
2. Ríos C, Herrero V. Producción científica latinoamericana y la ciencia mundial: una revisión bibliográfica (1989-2003). *Rev Interamericana de Bibliotecología*. 2005; 28(1):43-61.
3. Green BN, Johnson CD, Adams A. Writing narrative literature. *Reviews for peer-reviewed journals: Secrets of the Trade*. *J Sport Chiropr y Rehabil*. 2001; 15(1):5-19.
4. Rodríguez V. Conociendo sobre revisiones sistemáticas. *Rev Arch Col Med*. 2008; 1(2):64-68.
5. The Cochrane Collaboration. *The Systematic Review* [citado 2012 mayo 20]. Disponible en: www.cochrane.org/review/revstruc.html
6. The Joanna Briggs Institute (JBI). [citado 2012 mayo 20]. Disponible en: <http://connect.jbiconnectplus.org>

Aunque con estos resultados se demuestra la utilidad de los instrumentos para el estudio cuidadoso de piezas de investigación, es opcional que posteriormente se lleven a cabo otras pruebas de validez de criterio y confiabilidad a fin de precisar la estabilidad de los mismos. Este aspecto será presentado posteriormente por las autoras, que están haciendo uso del instrumento dentro del desarrollo del proyecto de investigación centrado en MA. En resumen, los instrumentos pueden ser tomados como base para evaluar diferentes tipos de estudios de abordaje cuantitativo. Se diferencian de otros porque no solo presentan el listado de características que deben evaluarse, sino que proponen una escala de puntuación para clasificar, según la calidad metodológica y el mérito científico de los artículos, la evidencia existente. Este aspecto puede ayudar a determinar qué literatura es útil para responder a problemas de la práctica profesional.

A futuro se espera fortalecer la práctica basada en evidencia y poder ofrecer al mundo de la investigación y de la práctica instrumentos con dotes psicométricas adicionales útiles para la selección y la lectura crítica de la literatura científica en las disciplinas de la salud.

Agradecimientos

Al Programa de Doctorado en Enfermería de la Universidad Nacional de Colombia; a Colciencias por el apoyo para adelantar estudios de doctorado; al Departamento de Metodología de la Investigación de la Facultad de Psicología de la Universidad de Murcia (España), por el apoyo recibido durante la realización de este estudio.

7. Junquera LM, Baladron J, Albertos JM, Olay S. Medicina basada en la evidencia (MBE): ventajas. *Rev Esp Cirug Oral y Maxilofac.* 2003;(25)5:265-272. Disponible en: <http://dx.doi.org/10.4321/S1130-05582003000500003>.
8. Lectura crítica. [Citado 2012 mayo 20]. Disponible en: www.redcaspe.org/caspe/lectura-critica
9. Bobenrieth MA. Normas para la revisión de artículos originales en ciencias de la salud. *Revista Internacional de Psicología Clínica y de la Salud.* 2002; 2(1): 509-523.
10. Bobenrieth MA. Lectura crítica de artículos originales en Salud. *Medicina en Familia.* 2002; 2(1): 81-89.
11. Gómez M, Danglot C, Velásquez L. Bases para la revisión crítica de artículos médicos. *Revista Mexicana de Pediatría.* 2001; 68(4):152-159.
12. Des Jarlais DC, Lyles C, Crepaz C and the Trend Group. Improving the Reporting Quality of Nonrandomized Evaluations of Behavioral and Public Health Interventions: The Trend Statement. *American Journal of Public Health.* 2004; 94(3): 361-366.
13. Coello PA, Ratoeche R, Etxeterria A. El sistema Grade para la evaluación de la calidad de la evidencia y la graduación de la fuerza de las recomendaciones. *Guías Clínicas.* 2008 [citado 2012 agosto 25]; (42): 8. Disponible en: www.garrahan.gov.ar/index.php/docencia-e-investigacion/investigacion-tecnologica/170-sistema-grade
14. Berra S, Elorza JM, Estrada MD, Sánchez E. Instrumento para la lectura crítica y la evaluación de estudios epidemiológicos transversales. *Gac Sant.* 2008; 22(5): 492-497.
15. Botella J, Gambara H. *¿Qué es el meta-análisis?* Madrid: Editorial Biblioteca Nueva; 2002.
16. DeCs. Descriptores en Ciencias de la Salud [citado 2011 octubre 31]. Disponible en: <http://decs.bvs.br/cgi-bin/wxis1660.exe/decserver>
17. Critical Appraisal Skills Programme Spanish. Lectura crítica [citado 2011 octubre 22]. Disponible en: www.redcaspe.org/que-hacemos/herramientas
18. Amaya MCP. Creación de la estructura para investigación en enfermería. Área temática: Fenómenos en Enfermería. Inducción de posgrados en enfermería. Maestría y doctorado. Universidad Nacional de Colombia; 2008. Actualizada para asuntos de Docencia en Investigación; 2012.
19. Contreras A, Flórez IE, Herrera LA. Un instrumento para evaluar la adherencia: su validez facial y confiabilidad. *Avances en Enfermería.* 2008;(26)2:39.
20. Sánchez E, Echeverry J. Validación de Escalas de Medición en Salud. *Revista de Salud Pública.* 2004;(6)3: 302-318.
21. Hernández B, Fernández C, Baptista P. *Metodología de la Investigación*, 5ª ed. México: McGraw-Hill; 2010. p. 200-204.
22. Polit D, Hungler PB. *Investigación científica en Ciencias de la Salud.* México: McGraw-Hill; 2000.
23. Lamprea JA, Gómez C. Validez en la Evaluación de Escalas. *Revista Colombiana de Psiquiatría.* 2007; 36(2):340-348.
24. Tristán A. Modificación al modelo de Lawshe para el dictamen cuantitativo de un instrumento objetivo. *Avances en Medición.* 2008;(6):37-48.
25. López de Ullibarri Galparsoro I, Pita S. Medidas de concordancia: el índice de Kappa. *Cad Aten Primaria.* 1999 [Visitada 2013 ago 10]; 6: 169. Disponible en: www.fisterra.com
26. Botella J, Gambara H. *¿Qué es el Meta-análisis?* 2ª ed. España: Biblioteca Nueva; 2002; p. 102-111.
27. Pita Fernández S, Pértegas Díaz S. La fiabilidad de las mediciones clínicas: el análisis de concordancia para variables numéricas [citado 2013 agosto 10]. Disponible en: www.fisterra.com
28. Torres JJ, Perera VH. Cálculo de la fiabilidad y concordancia entre codificadores de un sistema de categorías para el estudio del foro online en e-learning. *Revista de Investigación Educativa.* 2009; 27(1):89-103.
29. Montero I, León OG. Clasificación y descripción de las metodologías de investigación en psicología. *Revista Internacional de Psicología Clínica y de la Salud.* 2002;(2)3:503-508.

30. Carretero H, Pérez C. Normas para el desarrollo de estudios instrumentales. *Revista Internacional de Psicología Clínica y de la Salud*. 2005;(5)3: 521-551.
31. DRAE. Definición del término ítem. [citado 2013 agosto 10]. Disponible en: <http://buscon.rae.es/draeI/SrvltGUIBusUsual>
32. Centro Virtual Cervantes. Definición del término ítem. [citado 2013 agosto 10]. Disponible en: http://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/item.htm
33. LoBiondo-Wood G, Haber J. *Nursing research: Methods, critical, appraisal, and utilization*. 4th ed. México: Ed Elsevier 1998. p. 206-207.
34. Gutiérrez C et al. Validez y confiabilidad de la versión en español del instrumento "Escala de medición del proceso de afrontamiento y adaptación" de Callista Roy. *Aquichan*. 2007; 7(1):54-62.