



ORIGINAL

Validez de contenido por juicio de expertos de un instrumento para medir percepciones físico-emocionales en la práctica de disección anatómica



Martha Inés Bernal-García^{a,*}, David Ricardo Salamanca Jiménez^a,
Norton Perez Gutiérrez^b y Mónica Paola Quemba Mesa^c

^a Programa de Medicina, Grupo de Investigación HYGEA, Facultad de Ciencias de la Salud, Universidad de Boyacá, Tunja, Colombia

^b Programa de Medicina, Grupo de Investigación Villavicencio, Universidad Cooperativa de Colombia, Corporación Clínica Universidad Cooperativa de Colombia, Villavicencio, Colombia

^c Programa de Enfermería, Grupo de Investigación HYGEA, Facultad de Ciencias de la Salud, Universidad de Boyacá, Tunja, Colombia

Recibido el 5 de mayo de 2018; aceptado el 23 de agosto de 2018

Disponible en Internet el 30 de octubre de 2018

PALABRAS CLAVE

Validez de contenido;
Juicio de expertos;
Coeficiente kappa
de Fleiss;
Anatomía;
Disección;
Educación Médica

Resumen

Introducción: La presente investigación tuvo como objetivo realizar validez de contenido a partir del juicio de expertos de un instrumento de medición que indaga 4 dimensiones: reacciones físicas, consecuencias, perturbaciones y métodos de afrontamiento de los estudiantes de medicina ante la práctica de disección anatómica.

Material y método: Estudio psicométrico de validez de contenido, en el que participaron 9 expertos que evaluaron cada ítem con base en las características de suficiencia, claridad, coherencia, relevancia y pertinencia, al igual que estimaciones relacionadas con la congruencia de los ítems, amplitud de contenido, redacción, claridad y pertinencia del instrumento original. Fueron tenidas en cuenta las observaciones cualitativas de los expertos. Se determinó el grado de acuerdo entre expertos con el coeficiente kappa de Fleiss. Se aplicó prueba de pretest, para medir el grado de comprensibilidad.

Resultados: Globalmente se obtuvo fuerza de concordancia en cada dimensión, como casi perfecta y por parejas de expertos, entre moderada y casi perfecta. Para las características, la pertinencia fue la más alta (0,8443) y en la significación estadística, la suficiencia ($p=0,0268$). Se ajustó la estructura del instrumento, sin afectar la validez de contenido. De los 39 ítems originales, se eliminaron 16 ítems y se conservaron 22, de los que 6 se corrigieron en redacción. El grado de comprensibilidad fue de rango alto para la nueva versión del instrumento.

* Autor para correspondencia.

Correo electrónico: mibernal@uniboyaca.edu.co (M.I. Bernal-García).

KEYWORDS

Content validity;
Experts judgment;
Fleiss kappa
coefficient;
Anatomy;
Dissection;
Medical education

Conclusión: El instrumento validado podrá ser explorado y aplicado para favorecer la comparación de poblaciones de diferentes facultades en este y otros países que hablen el mismo idioma español.

© 2018 Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Content validity by expert judgment of an instrument to measure physico-emotional perceptions in anatomical dissection practice

Abstract

Introduction: The objective of the present investigation was to carry out content validity based on expert judgment of a measurement instrument that investigates 4 dimensions: physical reactions, consequences, disturbances and coping methods of students of medicine before the practice of anatomical dissection.

Material and method: Psychometric study of content validity, in which 9 experts participated who evaluated each item based on the characteristics of sufficiency, clarity, coherence, relevance and pertinence, as well as estimates related to the congruence of the items, breadth of content, writing, clarity and relevance of the original instrument. The qualitative observations of the experts were taken into account. The degree of agreement between experts was determined with the Fleiss kappa coefficient. A preliminary test was applied to measure the degree of comprehensibility.

Results: Globally, almost perfect strength of agreement was obtained in each dimension, and in pairs of experts, between moderate and almost perfect. For the characteristics, the relevance was the highest (0.8443) and in the statistical significance, the sufficiency ($p = .0268$). The structure of the instrument was adjusted, without affecting the content validity. Of the 39 original items, 16 items were eliminated, and 22 items were retained, of which 6 were corrected in writing. The degree of comprehensibility was of high rank for the new version of the instrument.

Conclusion: The validated instrument can be explored and applied to favor the comparison of populations of different faculties in this and other countries that speak the same Spanish language.

© 2018 Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

En el proceso de validación de instrumentos de investigación, como tema de interés en el área de la salud, es necesario disponer de escalas de medición para valorar los atributos subjetivos que integran constructos y dimensiones complejas¹ como lo son las físicas, psicológicas o sociales, entre otras, cuyo diseño admite reunir de forma válida y confiable la percepción del sujeto en diferentes momentos²⁻⁴ conducente a la generación de nuevo conocimiento. En ese sentido, se ha comunicado la descripción de los estándares que han venido evolucionando a través del tiempo para los conceptos de validez, últimamente basada en la evidencia y el marco de validación centrado en el contenido, los procesos de respuesta, la estructura interna, las relaciones con otras variables y las consecuencias del uso⁵. De estos, el concepto de validez de contenido ha sido objeto de múltiples transformaciones desde sus orígenes. Sin embargo, han estado más bien focalizadas en otorgarle importancia como fuente de evidencia de validez que en su definición operativa, que ha permanecido esencialmente estable, por lo que se hace necesaria para realizar interpretaciones de las puntuaciones a los ítems del instrumento

de medida, instrucciones para su administración y los criterios para su corrección y puntuación^{6,7}. Existen variados métodos para estimar la validez de contenido, entre ellos el basado en el juicio de expertos, que se caracteriza por contar con un número de expertos que bien proponen los ítems o dimensiones que deben conformar el constructo de interés o evalúan los diferentes ítems en función de su relevancia y representatividad⁷.

El incremento de la tasa poblacional y la necesidad de investigación multicultural conducen a la generación de datos fiables y válidos a partir de medidas de valoración cruzada como fuentes de estudio reproducibles en la población humana^{8,9}. No obstante, es limitada la información encontrada en la literatura que explique el proceso de validación de contenido de instrumentos para investigaciones específicas en el campo de la Medicina, aplicados en áreas básicas de formación profesional, como la asignatura de morfología humana, que lleva inmerso en su contenido programático diferentes técnicas de estudio y sesiones prácticas, entre la que se destaca la disección anatómica¹⁰, como método de enseñanza más antiguo.

Diversos estudios han demostrado que la disección estimula positivamente a los estudiantes por conocer en detalle

el cuerpo humano, promueve el trabajo en equipo, desarrolla habilidades prácticas, ayuda a la comprensión de la variabilidad anatómica, fomenta el aprendizaje con la interacción entre compañeros y ayuda a apreciar la vida humana a través de la comprensión de la muerte^{11,12}. Sin embargo, el primer acercamiento puede generar reacciones fisiológicas y psicológicas, no contempladas en varias facultades de medicina, lo cual podría tener repercusión en el proceso de aprendizaje, así como en la futura relación médico paciente desde la concepción del cuerpo humano como ente físico-espiritual¹³. Por ello, es necesario valorar objetivamente la percepción de los estudiantes frente a la práctica de disección anatómica, teniendo como pilar un instrumento de recolección de datos funcional, según la naturaleza del contenido, que simplifique la experiencia vivida por el estudiante en el anfiteatro.

Así, el presente trabajo de investigación tiene como propósito realizar proceso de validación de contenido a partir del juicio de expertos en el área de anatomía, de un instrumento de medición que indaga 4 dimensiones principales: reacciones físicas, consecuencias, perturbaciones y métodos de afrontamiento de los estudiantes de medicina ante la práctica de disección anatómica.

Método

Estudio descriptivo, psicométrico de validez de contenido por medio de juicio de expertos, realizado en la Universidad de Boyacá (Colombia).

Participantes

La población fue conformada por profesionales con experiencia de 10 o más años en la docencia e investigación en el campo de la anatomía humana con formación profesional posgradual de maestría o doctorado. La muestra, por intencionalidad, incluyó 9 expertos de diferentes Facultades de Medicina pertenecientes a Universidades públicas y privadas de Colombia, que aceptaron voluntariamente participar. Los criterios de exclusión utilizados fueron revocatoria del consentimiento informado e incumplimiento en la entrega del material evaluado en el plazo fijado.

Se contó con la participación de 2 asesores estadísticos con maestría en este campo.

Instrumentos

La validación fue aplicada sobre los test diseñados originalmente por Maribel Miguel y colaboradores para el estudio desarrollado durante 6 años, con 350 estudiantes en la población de primero de medicina de la Facultad de Medicina-Campus de Bellvitge de la Universidad de Barcelona, cuyo objetivo principal fue dar a conocer las actitudes, reacciones y medios de adaptación de los estudiantes, mientras realizaban las clases prácticas de Anatomía Humana del Aparato Locomotor¹⁴. El contenido se basó en la teoría Clásica de los Test, con ítems conceptuales para estudiar las reacciones de los estudiantes de primero de medicina ante la sala de disección, originalmente constituido por 39 ítems agrupados en

4 dimensiones: *las reacciones físicas* con 9 ítems, *consecuencias* con 7 ítems, *perturbaciones* con 9 ítems y *métodos de afrontamiento* con 14 ítems, percibidas en el anfiteatro (sala de disección) durante las prácticas de disección anatómica, como variables descritas previamente por esos autores. Para efectos de la presente investigación, se obtuvo permiso escrito de la autora Maribel Miguel.

La evaluación de los test agrupados en 4 partes (dimensiones) se hizo a partir de la valoración de cada ítem, utilizando la plantilla Juicio de expertos, propuesta por Escobar-Pérez y Cuervo-Martínez¹⁵, cuya calificación establece 4 niveles¹: no cumple con el criterio², Bajo nivel³, Moderado⁴ y Alto nivel, con base en la definición de indicadores en correspondencia con las características a evaluar: suficiencia, claridad, coherencia, relevancia y pertinencia, considerando además las observaciones cualitativas de los expertos, para cada uno de los 39 ítems que conformaron el instrumento original. Finalmente, cada uno de los expertos diligenció y firmó una constancia de validación, donde evaluaron globalmente las apreciaciones relacionadas con la congruencia de los ítems, amplitud de contenido, redacción, claridad y pertinencia del instrumento original.

Procedimiento

Los datos recogidos fueron almacenados en el programa Microsoft Excel 2016 y procesados estadísticamente en el programa STATA versión 14.0. Consecutivamente, se determinó el grado de acuerdo entre los expertos con el coeficiente kappa de Fleiss, como estadístico de análisis, utilizado para evaluar la concordancia entre 3 o más evaluadores que juzgan independientemente criterios de medida, a través de un instrumento, que consiste en un determinado número de categorías de naturaleza ordinal^{16,17}. Aplicable a estudios interjuicio, como en este caso, que asignó rangos a los ítems, basados en los datos ordinales y grado de varianza de la suma de los rangos 1-4, obtenidos por los diferentes expertos. El mínimo valor asumido por el coeficiente es 0 y el máximo 1¹⁴. Para la interpretación de este coeficiente se tuvo en cuenta la escala establecida por Landis y Koch (1977)¹⁸ que expresa cualitativamente la fuerza de concordancia entre los evaluadores (originalmente en inglés), detallada en la [tabla 1](#).

Para el análisis de las apreciaciones cualitativas de las jueces consignadas en la constancia de validación, sobre la

Tabla 1 Valoración del coeficiente kappa (Landis y Koch, 1977)

Coeficiente kappa de Fleiss	Fuerza de concordancia
0,00	Pobre (<i>Poor</i>)
0,1-0,20	Leve (<i>Slight</i>)
0,21-0,40	Aceptable (<i>Fair</i>)
0,41-0,60	Moderada (<i>Moderate</i>)
0,61-0,80	Considerable (<i>Substantial</i>)
0,81-1,0	Casi perfecta (<i>Almost perfect</i>)

Fuente: Landis y Koch¹⁸.

Tabla 2 Fuerza de concordancia entre evaluadores para las dimensiones del instrumento original

Dimensiones	Coefficiente kappa de Fleiss	Fuerza de concordancia Interpretación Landis y Koch, 1977
Sociodemográficas	0,970	Casi perfecta (<i>Almost perfect</i>)
Reacciones físicas	0,788	Considerable (<i>Substantial</i>)
Consecuencias	0,894	Casi perfecta (<i>Almost perfect</i>)
Perturbaciones	0,885	Casi perfecta (<i>Almost perfect</i>)
Métodos de afrontamiento	0,836	Casi perfecta (<i>Almost perfect</i>)

congruencia de los ítems, amplitud de contenido, redacción, claridad y pertinencia, se utilizó una escala de valoración tipo Likert, atribuyendo categorías y puntos de calificación, deficiente¹, aceptable², bueno³ y excelente⁴, según su consideración. Se estimaron adecuados los ítems que cumplieron con la totalidad de los requisitos establecidos, parcialmente adecuados los ítems que exigieron algunos cambios e inadecuados los que expresaron total incongruencia en relación con los criterios expresados. Se solicitó a los jueces la devolución del material tras el análisis y juicio final, vía electrónica a los investigadores, dentro del plazo establecido de 30 días. En los casos donde se consideró inadecuada, los motivos fueron explicados, las sugerencias descritas y los contenidos rehechos y mejorados.

Posteriormente, el instrumento obtenido de la validación de contenido por juicio de expertos fue sometido a la prueba de pretest estándar, con una codificación de respuestas, en función del número de opciones dicotómicas *sí/no*, sobre la interpretación de los ítems, en cada una de las dimensiones, para medir aplicabilidad/viabilidad; en una muestra aleatoria de 110 estudiantes, que cursaban la asignatura de morfología, en el programa de Medicina de la Universidad de Boyacá (Colombia), de los que se obtuvo individualmente el consentimiento informado, aceptado y firmado voluntariamente. Con estos datos se determinó el porcentaje de comprensión para todos los ítems y el grado de comprensibilidad en los siguientes rangos: Alta comprensibilidad igual o superior al 85%; mediana comprensibilidad del 80-85%; y Baja comprensibilidad, menor del 80%, de la nueva versión obtenida.

La presente investigación tuvo en cuenta los principios éticos de la declaración de Helsinki, de la Asociación Médica Mundial, y la resolución 008430 de 1993 del Ministerio de Salud y Protección social de Colombia, en la que todos los participantes tuvieron pleno conocimiento de los propósitos, la participación fue voluntaria y se aceptó con la firma del consentimiento informado. Asimismo, contó con el aval del Comité de investigación de la Facultad de Ciencias de la

Salud y del Comité de ética de la Universidad de Boyacá (Colombia)

Resultados

Validación de contenido por juicio de expertos

Sobre la evaluación del instrumento original, el cálculo del coeficiente kappa de Fleiss, tuvo en cuenta la proporción de posibles acuerdos que ocurrieron en cada dimensión, y estimó la magnitud en la fuerza de concordancia, interpretada como *casi perfecta* por el conjunto de los jueces, según se detalla en la [tabla 2](#).

De la misma manera se obtuvo magnitud en la fuerza de concordancia por parejas de expertos, de los cuales fue evidente la variabilidad entre moderada y casi perfecta, demostrando discordancia, especialmente en la dimensión de las reacciones físicas. Resultados detallados en la [tabla 3](#).

Asimismo, fueron estimadas las características del instrumento, para los indicadores categóricos de suficiencia, claridad, coherencia, relevancia y pertinencia, con base en la escala de medición ordinal; se encontró una fuerza de concordancia entre considerable y casi perfecta, siendo la característica de pertinencia la más alta (0,8443) según el grado de acuerdo global entre los jueces. En cuanto a la significación estadística de estas características, tuvo un comportamiento del 95% de seguridad llevando implícito una $p < 0,05$, en la que fue relevante la característica de suficiencia $p = 0,0268$. Así, el comportamiento estadístico que arroja coeficiente kappa de Fleiss y valor p como valor más informativo de cada característica es representado en la [tabla 4](#).

Posteriormente se valoraron las apreciaciones cualitativas globales del instrumento dentro de la constancia final de validación, con promedio (?) de puntuación asignada por los expertos como apreciación final y desviación estándar

Tabla 3 Concordancia por parejas entre los expertos

Dimensiones	Coeficiente kappa de Fleiss - Concordancia por parejas de expertos								
	1-9	2-8	3-7	4-6	5-4	6-3	7-2	8-1	9-5
Sociodemográficas	1	0,866	1	1	1	1	1	0,866	1
Reacciones físicas	0,722	0,863	1	0,572	0,694	0,766	1	0,669	0,805
Consecuencias	0,594	0,809	1	0,857	0,971	0,857	1	0,695	0,971
Perturbaciones	0,594	1	1	0,888	0,888	1	1	0,805	0,788
Métodos de afrontamiento	0,779	1	1	0,888	0,888	1	1	0,805	0,788

Tabla 4 Coeficiente de kappa de Fleiss y significación estadística de las características del instrumento original

Características	Coeficiente kappa de Fleiss	p
Suficiencia	0,804	0,026
Claridad	0,833	0,029
Coherencia	0,775	0,033
Relevancia	0,832	0,028
Pertinencia	0,844	0,031

(σ) medidas cuantitativamente en 5 dominios: Congruencia de ítem (? 3,6; con σ 0,5), Amplitud de contenido (? 3,3; con σ 0,7), Redacción de los ítems (? 3,3; con σ 0,7), Claridad y precisión(? 3,1; con σ 0,6), Pertinencia (? 3,6; con σ 0,5). Así, se obtuvo un resultado de calificación excelente en la valoración final, con puntuación promedio oscilante entre bueno³ y excelente⁴, mostrando que la concordancia entre los expertos estuvo homogéneamente de acuerdo, como comportamiento general, de las variables categóricas, según la estructura del instrumento de medida.

Con los resultados anteriormente descritos, y considerando las observaciones y recomendaciones emitidas cualitativamente por los expertos especialmente relacionadas con la claridad y precisión al conjunto de los 39 ítems contenidos en las dimensiones establecidas, fueron utilizadas para ajustar la estructura del instrumento de medida, por lo cual se eliminaron 17 ítems (2 de reacciones, 4 de consecuencias, 5 de perturbaciones y 6 de métodos de afrontamiento); de los 22 ítems que quedaron, 6 se corrigieron en redacción. Así, se obtuvo un instrumento de medida más claro, con menor número de ítems que permite disminuir el tiempo de aplicación y mejorar la objetividad para las opciones de respuesta.

La nueva versión del instrumento validado se organizó en 2 partes, teniendo en cuenta la temporalidad de su aplicación, la primera y última práctica; presenta un encabezado que otorga el título «Estudio de percepción en la práctica de disección-2 partes» versión corta, y abreviatura de identificación EPPD-2p, recoge datos sociodemográficos del participante, tiene de manera breve las indicaciones de diligenciamiento y detalla en cada dimensión los ítems, definiendo formato de numeración con nivel de

Tabla 5 Versión final del instrumento validado en contenido por juicio de expertos

Título del instrumento	Estudio de percepción en la práctica de disección-2 partes. Versión corta EPPD-2p		
Datos sociodemográficos del participante	A. Nombre B. Edad C. Sexo		
Indicaciones de diligenciamiento	Lea atentamente uno a uno los ítems que integra cada dimensión a evaluar, tenga en cuenta el momento de aplicación de este instrumento y señale con una X en el espacio lineado, según corresponda: primera práctica ____ última práctica _____, seleccione de manera objetiva y sincera, y asigne con una X la opción de respuesta que mejor se ajuste a su percepción		
Dimensiones	Ítem	Sí	No
1. Reacciones físicas ante la sala de disección (anfiteatro)	1a. Vértigo		
	1b. Náuseas		
	1c. Desmayo		
	1d. Palpitaciones		
	1e. Temblores		
	1f. Sudoración		
	1g. Abandono de la sala de disección (anfiteatro) si fuera posible		
2. Consecuencias	2a. Insomnio		
	2b. Pérdida del apetito		
	3c. Tener imágenes visuales recurrentes de los cadáveres		
3. Perturbaciones	3a. Olor de la sala de disección (anfiteatro)		
	3b. Ver todo el cadáver		
	3c. Ver una pieza anatómica		
	3d. Miedo a la infección		
4. Métodos de afrontamiento o evitación	4a. Pensar en otras cosas		
	4b. Relajación		
	4c. Pedir consejo a los compañeros		
	4d. Pedir consejo al profesor		
	4e. Usar tranquilizantes		
	4f. Evitar comer antes de entrar a la sala de disección (anfiteatro)		
	4g. Usar implementos de bioseguridad		
	4h. Evitar tocar el cadáver		

Tabla 6 Porcentaje de comprensión de las dimensiones y sus ítems en la versión final del instrumento validado

Dimensiones	Ítem	Porcentaje de comprensión %	
		Sí	No
Reacciones físicas ante la sala de disección (anfiteatro)	Vértigo	89,1	10,9
	Náuseas	100	0
	Desmayo	99,1	0,9
	Palpitaciones	90,0	10
	Temblores	97,3	2,7
	Sudoración	100	0
Consecuencia	Abandono de la sala de disección (anfiteatro) si fuera posible	90	10
	Insomnio	97,3	2,7
	Pérdida del apetito	97,3	2,7
	Tener imágenes visuales recurrentes de los cadáveres	90	10,0
Perturbaciones	Olor de la sala de disección (anfiteatro)	100	0
	Ver todo el cadáver	98,2	1,8
	Ver una pieza anatómica	100	0
	Miedo a la infección	90,9	9,1
Métodos de afrontamiento o evitación	Pensar en otras cosas	93,6	6,4
	Relajación	95,5	4,5
	Pedir consejo a los compañeros	96,4	3,6
	Pedir consejo al profesor	97,3	2,7
	Usar tranquilizantes	83,6	16,4
	Evitar comer antes de entrar a la sala de disección (anfiteatro)	92,7	7,3
	Usar implementos de bioseguridad	100	0
	Evitar tocar el cadáver	91,8	8,2

lista y escalas de valoración En la [tabla 5](#) se presenta la versión final del instrumento validado en la presente investigación.

Pretest

Según el número de casos procesados ($n = 110$), no se excluyó ninguno y se obtuvo como resultado el porcentaje de comprensión mediante respuesta dicotómica *sí o no*, asignada por los estudiantes en cada uno de los ítems que conformaron la estructura de la versión final instrumento validado por expertos. Con base en la distribución porcentual de comprensión obtenida para cada uno de los ítems ([tabla 6](#)), se determinó el grado de comprensibilidad en el rango alto (igual o superior al 85%), estableciendo así una comprensibilidad global esta nueva versión del instrumento de un 95%.

Discusión

Se presenta el proceso de validación de contenido por juicio de expertos en un instrumento que determina las percepciones físico-emocionales de los estudiantes de medicina frente a la práctica de disección anatómica en el anfiteatro. La fortaleza del estudio se relaciona con el proceso de adaptación, validación de las propiedades psicométricas y redacción del instrumento, el cual fue estructurado originalmente por investigadores españoles y validado culturalmente al contexto de la práctica académica colombiana, a partir de un

diseño metodológico que sintetiza paso por paso los distintos aspectos psicométricos, rigor y ayuda estadística^{19,20} que todo trabajo de investigación en medicina requiere para exponer resultados de impacto, principalmente para el campo de la educación en salud, dado que la metodología para adaptar instrumentos es poco conocida por profesionales sanitarios, lo que explica la existencia y uso indiscriminado de instrumentos adaptados de manera incompleta o traducciones simplemente literarias, por lo que requiere de un proceso dinámico y continuo que nunca acaba^{20,21}.

El presente trabajo de investigación es coherente con el interés en la realización y el uso apropiado de instrumentos de medición, partiendo del hecho de que la educación médica en las ciencias básicas constituye el pilar fundamental de la formación profesional^{22,23}. Evaluar las características de los ítems y preguntas es un criterio esencial para determinar la calidad en la medición de un instrumento²⁴, el cual es válido cuando es pertinente en la cultura donde se pretende aplicar; así, la adaptación de una escala internacional es muestra de la necesidad de indagar objetiva y comprensiblemente un sinnúmero de variables, que permiten cumplir los estándares estadísticos que fundamentan su aplicación²⁵.

Sin embargo, es previsible la aparición de dificultades inherentes al proceso validación de contenido, como la definición del número de expertos, el cual difiere considerablemente al consultar diferentes autores, dado que no hay acuerdo unánime para su determinación^{26,27}. Algunos señalan que deben ser 3 como mínimo²⁸, otros, fundamentados en la aplicación de variables estadísticas, concluyen

que la cantidad ideal varía entre 15 y 25 expertos^{29,30}; sin embargo, varios señalan que el número dependerá de los objetivos del estudio, con un rango entre 7 y 30 expertos^{31,32}. En ese sentido, el presente estudio seleccionó 9 expertos, atendiendo a los objetivos planteados, valor que se encuentra relacionado con la aproximación numérica citada por diversos autores.

La presente investigación convocó a expertos, basados en los criterios publicados especialmente enfocados a un proceso fundamental para asegurar asertividad en la evaluación; algunos de esos criterios son referidos a la experiencia del juez en la emisión de juicios y toma de decisiones, nivel de reputación en la comunidad académica y científica, disponibilidad y motivación para participar, mantener un importante grado de imparcialidad y cumplimiento¹⁵, estar capacitado en las técnicas de clasificación de preguntas para validez de contenido³³. Así, se buscaron expertos con el perfil descrito, para evitar sesgos de contenido en el análisis de datos.

Para el procesamiento analítico es importante definir una herramienta informática que determine el valor estadístico de la prueba. Clásicamente, este criterio ha sido abordado con el coeficiente kappa, originalmente propuesto por Cohen (1960) para evaluar el grado de concordancia entre 2 evaluadores³⁴; sin embargo, en el presente trabajo se aplicó la versión generalizada conocida como índice kappa de Fleiss, por tratarse de múltiples expertos (más de 3), basado en el concepto de noción de acuerdo entre diversos pares que incrementan la veracidad de los resultados¹⁶. Paralelamente, se realizó un análisis individualizado a cada ítem, empleando 4 criterios por indicador (suficiencia, claridad, coherencia y relevancia)³⁵ donde este coeficiente aseguró el rigor estadístico que soporta su aplicabilidad. Además, el análisis cualitativo de los ítems derivado de las observaciones de los expertos relacionadas con la congruencia, amplitud de contenido, redacción de los ítems, claridad y precisión, y pertinencia, permitió precisar razonamientos para corregir, conservar o eliminar ítems sin poner en riesgo la validez de contenido, cumpliendo con el proceso requerido para el estudio.

Utilizar un instrumento validado favorece el análisis de hipótesis aplicadas al campo de la educación³⁶ permitiendo acciones oportunas y eficaces para mejorar el proceso de aprendizaje en materias de alta dificultad para los estudiantes, como es el caso de la anatomía (morfología) humana. Ahora, tras haber completado el primer peldaño de la investigación, se presentará en una segunda comunicación, en conjunto con otras universidades colombianas, la percepción del alumnado ante la práctica de disección anatómica como actividad imprescindible en el proceso de formación médica.

Una medición realizada debe valorar apreciaciones teóricas y empíricas, por tanto, la medición se enfoca en la relación entre indicadores y conceptos no medibles directamente, como lo son las percepciones de una actividad académica, que establece la inferencia requerida al evaluar la aplicabilidad empírica de las proposiciones teóricas³⁷, obedeciendo con ello a las necesidades de la educación superior y expectativas de formación de futuros profesionales en medicina

Se hace ineludible realizar investigaciones similares que fortalezcan el campo de la educación básica en anatomía humana³⁸⁻⁴¹. Diversos autores convergen a la idea de que los estudiantes llegan a la práctica de disección con incertidumbre y múltiples sensaciones de ansiedad, precipitando la aparición de reacciones físicas y emocionales que podrían dificultar su proceso de aprendizaje. Tener conocimiento de estos factores y sensibilizar al alumno antes del comienzo de la actividad en el anfiteatro generaría beneficios mentales y emocionales necesarios para hacer su trabajo con entusiasmo y confianza.

Por lo anterior, es pertinente cuestionar las consecuencias vividas por los estudiantes en el anfiteatro, ya que la disección se ha mantenido como una herramienta de enseñanza primaria en la historia de la medicina y, como tal, se debe optimizar el recurso pedagógico que integra su reproducibilidad. La enseñanza basada en la disección es referida como esencial hasta la fecha. Su utilidad también se refleja en la percepción de los estudiantes, opinando que la disección les proporciona una base fundamental para el desarrollo de habilidades clínicas. Los investigadores incluso han sugerido que puede restablecerse la disección como el método central de enseñanza de la anatomía para garantizar la práctica médica segura^{42,43}.

Así, al haber cumplido con el objetivo primario de validar un instrumento de medición, consolidado y adaptado, podrá ser explorado y aplicado en diferentes comunidades académicas para determinar la percepción ante la práctica de disección, con nuevas pruebas psicométricas que avalen mayor validez y favorezcan la comparación de poblaciones de diferentes facultades en este y otros países que hablen el mismo idioma español, y establecer acciones oportunas que faciliten el proceso de aprendizaje orientado a las actividades prácticas propias de la anatomía macroscópica en el anfiteatro y otras salas de disección.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Bibliografía

1. Luján-Tangarife JA, Cardona-Arias JA. Construction and validation of measurement scales in health: A review of psychometric properties. *Arch Med*. 2015;11:1-10.
2. Montero-Rojas E. Referentes conceptuales y metodológicos sobre la noción moderna de validez de instrumentos de medición: implicaciones para el caso de personas con necesidades educativas especiales. *AP*. 2013;27:113-28.
3. Ahmed R, McCarroll M, Schwartz A, Gothard D, Atkinson S, Hughes P, et al. Development, validation, and implementation of a medical judgment metric. *Medical Decision Making Policy & Practice (MDM P&P)*. 2017;2:1-9.
4. Sánchez R, Gómez C. Conceptos básicos sobre la validación de escalas. *Rev Col Psiquiatría*. 1998;27:121-30.
5. Sireci S, Padilla J. Validating assessments: Introduction to the special section. *Psicothema*. 2014;26:97-9.
6. Pedrosa I, Suárez-Álvarez, García-Cueto E. Evidencias sobre la validez de contenido: avances teóricos y métodos para su estimación. *Acción Psicológica*. 2013;10:3-20.

7. Abad FJ, Olea J, Ponsoda V, García C. *Medición en ciencias sociales y de la salud*. Madrid, España: Editorial Síntesis; 2011.
8. Sousa VD, Zauszniewski JA, Mendes IAC, Zanetti ML. Cross-cultural equivalence and psychometric properties of the Portuguese version of the Depressive Cognition Scale. *J Nurs Meas*. 2005;13:87–99.
9. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *J Eval Clin Pract*. 2011;17:268–74.
10. Vivekananda-Schmidt P, Lewis M, Hassell AB, Coady D, Walker D, Kay L, et al. Validation of MSAT: An instrument to measure medical students' self-assessed confidence in musculoskeletal examination skills. *Med Educ*. 2007;41:402–10.
11. Elizondo R, Guzman S, García M. Dissection as a teaching tool: Past, present, and future. *Anat Rec B New Anat*. 2005;285:11–5.
12. Montemayor BG. El significado de la práctica de disección para los estudiantes de medicina. *Int J Morphol*. 2006;24:575–80.
13. Jagua-Gualdrón A, Urrego-Mendoza DZ. Actitudes de los estudiantes colombianos de medicina hacia la práctica de la disección en anatomía y su relación con el puntaje en la escala de empatía médica de Jefferson. *Rev Fac Med*. 2011;59:10–20.
14. Miguel M, Porta N, Ortiz J, Martínez A. Human anatomy: reactions of first year medical students to the dissection room. *Educación Médica*. 2007;10:105–13.
15. Escobar-Pérez J, Cuervo-Martínez Á. Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Av En Medición*. 2008;6:27–36.
16. McHugh ML. Interrater reliability: The kappa statistic. *Biochem Medica*. 2012;22:276–82.
17. Falotico R, Quatto P. Fleiss' kappa statistic without paradoxes. *Qual Quant*. 2015;49:463–70.
18. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–74.
19. Argimon JM, Jiménez J. *Validación de cuestionarios. Métodos de investigación clínica y epidemiológica*. Madrid: Elsevier; 2004.
20. Carvajal A, Centeno C, Watson R, Martínez M, Sanz Rubiales Á. ¿Cómo validar un instrumento de medida de la salud? *An Sist Sanit Navar*. 2011;34:63–72.
21. Abad F, Garrido J, Olea J, Ponsoda V. *Introducción a la Psicometría: Teoría Clásica de los Tests y Teoría de la Respuesta al Ítem*. Madrid: Universidad Autónoma de Madrid; 2006. p. 61–118.
22. Lamprea J, Gómez C. Validez en la evaluación de escalas. *Rev Colomb Psiquiat*. 2007;36:340–8.
23. Sánchez R, Echeverry J. Validación de escalas de medición en salud. *Rev Salud Pública*. 2004;6:302–18.
24. Polit DF, Beck CT. *Nursing research: principles and methods*. 7th ed. Philadelphia: Lippincott Williams & Wilkins; 2003.
25. Alarcón A, Muñoz S. Medición en salud: Algunas consideraciones metodológicas. *Rev Méd Chile*. 2008;136:125–30.
26. Powell C. The Delphi technique: Myths and realities. *J Adv Nurs*. 2003;41:376–82.
27. Cabero J, Llorente MC. La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información (TIC). *Eduweb*. 2013;7:11–22.
28. Delgado-Rico E, Carretero-Dios H, Ruch W. Content validity evidences in test development: An applied perspective. *Intern J Clin Health Psych*. 2012;12:449–60.
29. García L, Fernández SJ. Procedimiento de aplicación de trabajo creativo en grupo de expertos. *Rev Ingeniería Energética*. 2008;29:46–50.
30. Corral Y. Validez y confiabilidad de los instrumentos de Investigación para la recolección de datos. *Rev FCE*. 2009;19:229–47.
31. Varela Ruiz M, Diaz-Bravo L, García-Durán R. Descripción y usos de la técnica Delphi en investigaciones en áreas de la salud. *Inv Ed Med*. 2012;1:90–5.
32. Urrutia M, Barrios S, Gutiérrez Ñ, Mayorga M. Métodos óptimos para determinar validez de contenido. *Educ Med Super*. 2014;28:547–58.
33. Garrido ME, Romero S, Ortega E, Zagalaz ML. Designing and validation of a questionnaire on parents for children in sport. *J Sport Health Res*. 2011;3:59–70.
34. Marasini D, Quatto P, Ripamonti E. Assessing the inter-rater agreement for ordinal data through weighted indexes. *Stat Methods Med Res*. 2016;25:2611–33.
35. Dorantes-Nova J, Hernández-Mosqueda J, Tobón-Tobón S. Judgment of experts for the validation of an instrument of measurement of burnout syndrome in teaching. *Ra Ximhai*. 2016;12:327–46.
36. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm*. 2008;65:2276–84.
37. Soriano AM. Design and validation of measurement instruments. *Ed Universidad Don Bosco*. 2014;8:19–40.
38. Getachew D. Reaction of medical students to experiences in dissection room. *Ethiop J Health Sci*. 2014;20:337–42.
39. Qamar K, Osama M. Role of dissection in light of students' perceptions. *J Pak Med Assoc*. 2014;64:1021–4.
40. Van-Wyk J, Rennie C. Learning anatomy through dissection: Perceptions of a diverse medical student cohort. *Int J Morphol*. 2015;33:89–95.
41. Bob M, Popescu CA, Suci SM, Buzoianu AD. First year medical students' attitude toward anatomical corpse dissection and its relationship with their personality. *Rom J Morphol Embryol*. 2015;56:321–4.
42. Chosh SK, Sanjib. Cadaveric dissection as an educational tool for anatomical sciences in the 21st Century. *Anat Sci Educ*. 2017;10:286–99.
43. Winkelmann A. Anatomical dissection as a teaching method in medical school: A review of the evidence. *Med Educ*. 2007;41:15–22.