

Métodos óptimos para determinar validez de contenido

Optimal method for content validity

MSc. Marcela Urrutia Egaña, MSC. Silvia Barrios Araya, Lic. Marina Gutiérrez Núñez, Lic. Magdalena Mayorga Camus

Pontificia Universidad Católica de Chile. Chile.

RESUMEN

Introducción: la validez es un criterio que permite comprobar la calidad de un instrumento, verificando si este evalúa lo que pretende medir; es de vital importancia al considerar que estos instrumentos son utilizados en momentos relevantes, como la selección de nuevo personal o la aprobación o reprobación de un grado académico.

Objetivo: conocer las metodologías involucradas en el proceso de validez de contenido.

Métodos: se realizó una búsqueda de artículos en base de datos y, a partir de 58 artículos, se seleccionaron 40.

Resultados: se halló que el comité de expertos y la entrevista cognitiva son las metodologías más utilizadas para validar contenido; la primera presenta el desafío de la selección de los jueces, por su parte la entrevista cognitiva conlleva un alto costo, sin embargo ambas son aconsejables.

Conclusión: para un mejor resultado de validez de contenido se sugiere utilizar más de una metodología, con el fin de aumentar la rigurosidad del proceso.

Palabras clave: validez de contenido, validación por expertos, protocolo de pensamiento en voz alta/pensamiento hablado, entrevista cognitiva.

ABSTRACT

Introduction: validity is a criterion for checking the quality of an instrument, to see whether it really evaluates what it is intended to measure. It is of vital importance to consider that these instruments are used at important moments,

such as the selection of new staff or approval/disapproval of an academic degree.
Objective: to know the methodologies involved in the process of content validity.
Methods: article search in databases, 40 out of 58 were selected.
Results: the expert committee and the cognitive interview methodologies were used to validate the content; the former poses the challenge of selecting judges whereas the latter carries high cost; however, both are desirable.
Conclusions: for best results, it is suggested that content validity be evaluated with more than one methodology, in order to increase the thoroughness of the process.

Key words: content validity, expert validation, protocol thinking aloud / spoken thought, cognitive interview.

INTRODUCCIÓN

La validez es un criterio de evaluación utilizado para determinar cuán importantes son la evidencia empírica y los fundamentos teóricos que respaldan un instrumento, examen o acción realizada.¹⁻³ También, es entendida como el grado en que un instrumento mide lo que pretende medir o que cumple con el objetivo para el cual fue construido,^{3,4-6} este criterio es fundamental para considerar válida una prueba.⁷ La validez junto a la fiabilidad determinan la calidad que posee un instrumento.^{4,5}

En la actualidad, esta se ha transformado en un elemento relevante dentro de la medición⁶ por el aumento de nuevos instrumentos utilizados en momentos cruciales, por ejemplo al seleccionar nuevo personal o al determinar la aprobación o reprobación de un grado académico,⁸ así, mismo, hay quienes señalan la necesidad de validar el contenido de los instrumentos ya existentes.⁹

El proceso de validación es dinámico y continuo^{10,11} y va adquiriendo mayor relevancia a medida que se profundiza en él. La *American Psychological Association* (APA), en 1954, señaló 4 tipos de validez: de contenido, predictiva, concurrente y de constructo.^{5,7,12} Sin embargo, otros autores la clasifican en validez de apariencia, de contenido, de criterio y de constructo.¹³

La validez de contenido se define como el juicio lógico sobre la correspondencia que existe entre el rasgo o la característica del aprendizaje del evaluado y lo que se incluye en la prueba o examen.¹⁴ Pretende determinar si los ítemes o preguntas propuestas reflejan el dominio de contenido (conocimientos, habilidades o destrezas) que se desea medir.^{2,4,15} Para ello se deben reunir evidencias sobre la calidad y la relevancia técnica del *test*; es fundamental que sea representativo del contenido mediante una fuente válida, como por ejemplo: literatura, población relevante o la opinión de expertos.^{11,12,16,17} Lo anterior asegura que la prueba incluya solo aquello que debe contener en su totalidad, es decir, la pertinencia del instrumento.¹⁸

Este tipo de validez puede considerar criterios internos y externos. Entre los criterios de validez interna están la calidad del contenido, la importancia curricular, cobertura del contenido, complejidad cognitiva, adecuación lingüística, habilidades complementarias y el valor o ponderación que se dará a cada ítem.^{9,14,19} Entre los

criterios de validez externa se encuentran: la equidad, la transferencia y generalización, la comparabilidad y la sensibilidad de la instrucción; estas repercuten tanto en los estudiantes, como en los profesores.^{14,19}

El objetivo de la presente revisión es conocer las metodologías involucradas en el proceso de validez de contenido. Esta necesidad surge a partir de la decisión de optar por un examen escrito de selección múltiple, que mide conocimientos y habilidades cognitivas, como modalidad para obtener el título profesional de enfermera o enfermera matrona en una escuela de salud de una universidad chilena. Este proceso se inicia en el año 2003 con la elaboración de preguntas y el análisis psicométrico de estas, sin embargo, se considera fundamental determinar la validez de contenido del instrumento utilizado.

Para el logro de este objetivo se realizó una búsqueda en diferentes bases de datos de la colección electrónica, disponibles en el sistema de multibúsqueda de la Universidad, utilizando las palabras clave: *validez de contenido, validación por expertos, protocolo de pensamiento en voz alta/pensamiento hablado*. Para la selección de las publicaciones, los criterios de inclusión utilizados fueron: artículos publicados desde el año 2002 en adelante; a texto completo, sin restricción de idioma, cabe destacar que se incorporó bibliografía de autores clásicos en la materia. Se encontraron 58 artículos, de los cuales se seleccionaron 40.

La información encontrada se organizó en torno a las 2 metodologías más utilizadas para validar contenido: comité de expertos y entrevista cognitiva.

DESARROLLO

Tipo de validez de contenido

Existen diversas metodologías que permiten determinar la validez de contenido de un *test* o instrumento, algunos autores plantean que dentro de ellas se encuentran los resultados del *test*, la opinión de los estudiantes, las entrevistas cognitivas y la evaluación por parte de expertos;² otros realizan análisis estadísticos con diversas fórmulas matemáticas, por ejemplo, utilizan fórmulas factoriales con ecuaciones estructurales, estos son menos comunes.²⁰ En las entrevistas cognitivas se obtienen datos cualitativos en los que se puede profundizar; a diferencia de la evaluación por expertos que busca determinar la habilidad que pretende medir las preguntas del examen.² Algunos expertos señalan que para validar el contenido de un instrumento son fundamentales: la revisión de investigaciones, de incidentes críticos, la observación directa del instrumento aplicado, el juicio de expertos y los objetivos instruccionales.²¹ Los métodos señalados frecuentemente en los artículos revisados son el comité de expertos y la entrevista cognitiva.¹⁷

Comité de expertos

Es una metodología que permite determinar la validez del instrumento por medio de un panel de jueces expertos para cada una de las áreas curriculares a considerar en el instrumento de evaluación, quienes deben analizar –como mínimo– la coherencia de los ítems con los objetivos de los cursos, la complejidad de los ítems y la habilidad cognitiva a evaluar.¹² Los jueces deben contar con

capacitación en las técnicas de clasificación de preguntas para validez de contenido.^{6,19,22} Esta metodología es la más usada para realizar la validación de contenido.⁷

Es por ello fundamental que previo a realizar esta validación se resuelvan 2 problemas: primero, determinar aquello que puede ser medido y segundo, determinar quiénes serán los expertos que validarán el instrumento. Para lo primero, es esencial que el autor haga una exhaustiva revisión bibliográfica sobre el tema, puede trabajar a su vez con grupos focales; este período es definido por algunos autores como etapa de desarrollo.¹⁷

Para lo segundo, si bien no existe un consenso que defina las características de un experto, es fundamental que este conozca sobre el área a investigar, ya sea al nivel académico y/o profesional y que, a su vez, conozca de áreas complementarias.¹⁵⁻²³ Sin embargo, otros autores son más enfáticos al momento de definir quién es experto y consideran como requisito, por ejemplo, que cuente con al menos 5 años de experiencia en el área.¹ Todo esto obliga a que la muestra sea intencionada.²³

Se deben definir las características del experto y, a la vez, determinar el número de ellos. *Delgado* y otros señalan que deben ser 3, como mínimo,⁶ mientras que *García* y *Fernández*, al aplicar variables estadísticas, concluyeron que la cantidad ideal varía entre 15 y 25 expertos;²⁴ sin embargo, *Varela* y otros, señalan que el número dependerá de los objetivos del estudio, con un rango entre 7 y 30 expertos.²⁵ Existen otros autores menos estrictos a la hora de determinar el número de expertos, ellos consideran la existencia de diversos factores, tales como: el área geográfica o la actividad laboral, entre otros. Además, señalan que es fundamental prever el número de expertos que no podrán participar o que desertarán durante el proceso.²⁶

Una vez que se decide cuáles serán los criterios para seleccionar los expertos, se les invita a participar en el proyecto; durante el mismo período se procede a elaborar una matriz de clasificación, con los que cada juez determinará el grado de validez de las preguntas.¹⁶

Para el proceso de elaboración de la matriz se utiliza la escala Likert de 3, 4 o 5 puntos en donde la evaluación de las posibles respuestas podrían clasificarse en diferentes tipos, por ejemplo: a) excelente, buena, regular y mala;¹⁵ b) esencial; útil; útil, pero no esencial ni necesaria. Lo anterior está en dependencia del tipo de matriz y de los objetivos específicos que se persigan. Además, otros estudios mencionan haber incorporado espacios en donde el experto puede entregar sus aportes y apreciaciones con respecto a cada pregunta. Posteriormente, se le entrega a cada experto –vía correo electrónico o presencialmente en una oficina facilitada por el investigador– la matriz de clasificación y el instrumento a evaluar.¹⁵

Una vez que se obtienen los resultados de los expertos se procede al análisis de los datos; la manera más común es medir la concordancia de la evaluación del ítem en revisión, reportada por cada uno de los expertos, se considera aceptable cuando supera el 80 %; aquellos que no alcanzan este porcentaje pueden ser modificados y sometidos a un nuevo proceso de validación o simplemente ser eliminados del instrumento.^{1,27}

Otros autores reportan utilizar la prueba estadística de Lashe (1975), para determinar el grado de acuerdo entre los jueces, observan un ratio de validez de contenido con valores entre -1 y +1. Cuando el valor es positivo indica que más de la mitad de los jueces se encuentran de acuerdo; por el contrario, si este es

negativo, significa que menos de la mitad de los expertos lo están. Una vez obtenidos los valores se procede a modificar o eliminar las preguntas o ítemes.^{4,22}

Para determinar la validez de contenido mediante expertos, se proponen siguientes las fases: a) realizar una definición del universo de observaciones admisibles; b) determinar quiénes son los expertos en el universo; c) exponer –por parte de los expertos– el juicio por medio de un procedimiento concreto y estructurado sobre la validez del contenido y d) elaborar un documento donde se resumen los datos recopilados anteriormente.^{22,28}

La literatura describe otras metodologías que pueden ser utilizadas en conjunto o individualmente. Entre ellas se encuentran:

- Modelo de Fehring: pretende explorar si el instrumento mide el concepto que quiere medir con la opinión de un grupo de expertos; es utilizado en el ámbito de la enfermería, por la Asociación Americana de Diagnósticos de Enfermería (NANDA, por su nombre en inglés), para analizar la validez de las intervenciones y los resultados. El método consta de las fases siguientes:

a) Se seleccionan los expertos, quienes determinan la pertinencia y relevancia del tema y las áreas a evaluar mediante una escala de Likert.

b) Se determinan los puntajes asignados por los jueces y la proporción de estos en cada una de las categorías de la escala con lo que se obtiene el índice de validez de contenido (IVC); este índice se logra mediante la suma de cada uno de los indicadores entregados por los expertos en cada uno de los ítemes, y, finalmente, se divide por el número total de expertos. Se promedian cada uno de estos índices particulares, se descartan aquellos cuyo promedio no superara el 0,8.

c) Se edita de manera definitiva el formato del texto, tomando en cuenta el valor del IVC, según el parámetro antes mencionado, se determinan aquellos ítems que conformarán el instrumento final y los que por su bajo valor IVC son considerados críticos y deben ser revisados.²⁹

Un ejemplo de uso específico de este modelo fue la adaptación llevada cabo por *Fehring* para realizar la validez de contenido de los diagnósticos de enfermería; en este caso el autor propone 7 características que debe cumplir un experto, las cuales se asocian a un puntaje según su importancia, se espera obtener al menos 5 de ellas para ser seleccionado como experto. La máxima puntuación la obtiene el grado de Doctor de Enfermería (4 puntos) y uno de los criterios para el mínimo de los puntajes (1 punto) es poseer práctica clínica de un año en el área de interés; es importante aclarar que los autores reconocen la dificultad que existe en algunos países por la falta de experticia de los profesionales.³⁰

- Metodología Q: fue introducida por *Thompson y Stephenson* en el año 1935, con el fin de identificar de manera cualitativa-cuantitativa patrones comunes de opinión de los expertos con respecto a una situación o temática.³¹ La metodología es llevada a cabo por medio del sistema de ordenamiento Q, el cual está dividido en etapas: la primera reúne a los expertos según lo aconsejado por *Waltz* (entre 25 y 70), quienes seleccionan y ordenan las preguntas según sus puntos de vista sobre el tema en estudio, además, se les entrega evidencia bibliográfica como respaldo. La segunda fase consiste en recopilar esta información, por cada uno de los expertos, según relevancia, que va a lo largo de un continuo, desde "muy de acuerdo" hasta "muy en desacuerdo"; finalmente, se realizan análisis estadísticos

para determinar la similitud de toda la información y las dimensiones del fenómeno.³⁰

- Método Delphi: permite obtener la opinión de un panel de expertos; es utilizado cuando existe escasa evidencia empírica, los datos son difusos o predominan factores de índole subjetiva, permite a los expertos expresarse de manera libre ya que las opiniones son confidenciales; a su vez, evita problemas como la escasa representatividad y el dominio de algunas personas sobre otras.³²

Durante el proceso participan 2 grupos, uno de ellos elabora las preguntas y diseña ejercicios, denominado grupo monitor y el segundo, compuesto por expertos, las analiza. El grupo monitor adquiere un rol fundamental ya que debe manejar los objetivos del estudio y, además, cumplir con una serie de requisitos, tales como: conocer a cabalidad la metodología Delphi, ser investigador académico del tema a estudiar y tener habilidades para las relaciones interpersonales. Las rondas suceden en completo anonimato, los expertos opinan y debaten sobre las opiniones de otros pares, realizan sus comentarios y vuelven a analizar sus propias ideas con la retroalimentación de los otros participantes. Finalmente, el grupo monitor genera un informe que resume el análisis de cada una de las respuestas y estrategias entregadas por los expertos. Es fundamental que el número de rondas sea acotado por el riesgo de abandono del proceso por parte de los expertos.³³

Este último es el más utilizado por su alto grado de fiabilidad, flexibilidad, dinamismo y validez (de contenido y otras); dentro de sus atributos se destacan: el anonimato de los participantes, la heterogeneidad de los expertos, la interacción y retroalimentación prolongada entre los participantes, este último atributo es una ventaja que no está presente en los otros métodos. Además, existe evidencia que señala que es un aporte en la seguridad de la decisión tomada, ya que esta responsabilidad es compartida por todos los participante.³⁴

Entrevista cognitiva

Esta metodología fue definida según *Armengol*, como un instrumento metodológico que implica que los participantes piensan en voz alta mientras llevan a cabo la actividad solicitada.³⁵ *Ericsson* y *Simon* (1993) introducen la variable tiempo, al estipular que si el relato ocurre durante la ejecución de la tarea se denominará concurrente y cuando esta sea recordada, retrospectiva.³⁶ Los relatos concurrentes están relacionados con la memoria a corto plazo del participante y no se altera la secuencia de su pensamiento; lo contrario ocurre en los relatos retrospectivos que utilizan la memoria a largo plazo, en los que la secuencia de su pensamiento se ve alterada, ya que debe buscar información que no utilizó durante la realización de la tarea.³⁶⁻³⁸ Al usar ambos relatos para una misma actividad se genera mayor información acerca del proceso de pensamiento.^{36,37}

Los reportes verbales concurrentes comienzan en conjunto con la tarea. No se solicita que describan o expliquen qué están haciendo; a diferencia de los reportes verbales retrospectivos en el cual el registro comienza una vez finalizada la acción; la instrucción que se da a los usuarios es que *piensen en voz alta*. Se le solicita a la persona que recuerde todo lo posible acerca de sus pensamientos durante la ejecución de la actividad.³⁷ De esta manera, el investigador podrá observar las estrategias cognitivas y metacognitivas que la persona utiliza al momento de realizar la actividad que se le solicita; dichas estrategias aumentan si se compara con la misma persona en condiciones de silencio.^{39,40}

Los pensamientos hablados, del tipo que sean, deben ser grabados para posteriormente transcribirlos *verbatim* y analizarlos con la ayuda de categorías

preestablecidas.⁴⁰ No obstante, estos análisis deben ser realizados por más de una persona para evitar sesgos.³⁹

Esta metodología ha sido utilizada de manera amplia por diferentes disciplinas para analizar actividades de resolución de problemas, su origen es el área de la psicología cognitiva.³⁹ Entre ellas destacan la lingüística, educación, física, química y matemáticas.³⁷

Algunos autores refieren que esta práctica tuvo su origen en el método de la introspección⁴¹ a diferencia de otros que afirman que se inició cerca de los años 70, a partir del uso de los protocolos de pensamiento en voz alta, propuestos por *Emig*, para estudiar el acto de la composición de nativos hablantes de inglés.⁴⁰ Posteriormente, en la década de los 80, *Flower* y *Hayes* utilizan este protocolo para fundamentar su *Modelo cognitivo del proceso de composición* para hablantes de inglés, como segunda lengua y como lengua extranjera, lo que permite su masificación.⁴⁰

Durante la década de los 60, la psicología cognitiva refería que los procesos mentales y las habilidades generativo-creativas eran analizables.⁴¹ Además, junto con esto se discute a cerca de la explotación de los métodos empíricos de retrospección e introspección para analizar dichos procesos, los cuales habían sido rechazados por los investigadores conductistas. Lo anterior hace referencia a la necesidad de establecer un marco teórico que pudiera estudiar los procesos mentales de manera fiable³⁹ ya que las críticas a los reportes verbales señalan que es complejo para las personas verbalizar su pensamiento sin justificarlo, ni racionalizarlo.³⁷

Ese desafío fue asumido por *Ericson* y *Simon*, quienes gracias a la publicación de su influyente artículo *Verbal Reports as Data*, generaron un marco teórico explícito y riguroso el cual fue utilizado para fundamentar su teoría de procesamiento de información. Junto con esto, propusieron "analizar la validez de los reportes verbales como fuentes de datos, a partir de un modelo cognitivo basado en representaciones y procesos."³⁶ Según ambos autores, el pensar en voz alta no modifica el proceso cognitivo mientras el usuario realiza la tarea, sin embargo, puede utilizar más tiempo en concluirla ya que se enlentece su pensamiento.³⁶

Estos mismos autores analizados por *Armengol* plantean que, según las instrucciones dadas para el reporte del pensamiento hablado, se generan 3 niveles de informes orales; por esto, aconsejan planificar con cuidado la metodología de trabajo para obtener el informe deseado.³⁵

- Informes de nivel 1: provienen de la vocalización directa del pensamiento disponible en la memoria a corto plazo o del trabajo del participante (*short termmemory*) tras la instrucción simple de "pensar en voz alta continuamente y solo dejar de hacerlo cuando hayan terminado de realizar la actividad indicada" (*talkaloud*).

- Informes de nivel 2: reflejan descripciones o explicaciones sobre el contenido del pensamiento presente en la atención de la persona, sin intención de incitar a la búsqueda de información en la memoria a largo plazo (*longtermmemory*). Este requiere instrucciones que solicitan detalles complementarios, por ejemplo "articular todo lo que pasa por la mente y no dejar de hablar."⁴⁰ Esta instrucción ocasiona que la tarea se enlentezca, sin afectar el proceso cognitivo.

- Informes de nivel 3: activan los procesos intermedios para localizar información que no está presente en la memoria de trabajo y que necesita cierto grado de

introspección. Las instrucciones se orientan al control consciente por parte del usuario y pueden afectar tanto el curso como la estructura del proceso cognitivo en relación con el nivel de introspección solicitado.⁴⁰

En caso de que la persona deje de hablar, se sugiere decirle "siga hablando", en vez de "¿puede decirme que es lo que está pensando ahora?" ya que esta frase solo generará descripciones, justificaciones y explicaciones.³⁷

Para facilitar este proceso, *Ericson y Simon*, formulan un manual de procedimiento para recoger los datos, considerando que los procedimientos escogidos, la relación entre la información solicitada y la secuencia de los datos obtenidos, condicionan la veracidad de los datos.³⁶

Además, propusieron complementar los protocolos de pensamiento en voz alta con procedimiento de evocación (*recall*). Esto consiste en utilizar material relacionado con la tarea (el video de la grabación de la sesión y el documento escrito que se generó) para evitar explicaciones erróneas por parte de los usuarios cuando se les pregunte la razón de su pensamiento. Esto se debe realizar inmediatamente después que la persona termine su tarea a fin de "suministrar la información ausente de los protocolos a causa de los procesos automatizados y de los períodos de silencio."⁴⁰

De Brito reúne algunas recomendaciones propuestas por diversos autores, para que la metodología de la entrevista cognitiva sea utilizada correctamente, entre ellas: a) establecer ciertas categorías de análisis en dependencia del objetivo que tenga la investigación; b) determinar características de los participantes considerando los conocimientos previos, la capacidad de la memoria de trabajo, destrezas personales y motivación; c) definir el número de los participantes teniendo presente que los investigadores pueden analizar los datos de mejor manera cuando el número de personas es limitado; d) seleccionar el método de análisis; e) definir instrucciones para los participantes; este punto es controversial ya que algunos autores plantean que es necesario que las personas tengan conocimiento y experiencia en lo que significan las entrevistas cognitivas, mientras que otros postulan lo opuesto; f) finalmente, transcribir y analizar la información obtenida; es recomendable considerar el lenguaje no verbal del participante⁴¹

En la literatura se mencionan algunas desventajas de esta metodología, como por ejemplo, la dificultad para usarla con personas con alteraciones lingüísticas, no obstante, lo anterior puede verse facilitado si se le indica una tarea concreta.³⁷ También, se menciona que el rol de investigador en este tipo de protocolos orales es primordial ya que si este investigador guía sobremanera la tarea –realizando múltiples gestos y comentarios del contenido– podría verse alterada la validez del relato. Si el experto no cuida constantemente estos factores los posteriores análisis serán erróneos. Finalmente, el juicio y subjetividad de los investigadores es crucial para los resultados ya que en ocasiones los relatos pueden ser incompletos, irrelevantes o insuficientes.³⁹

Por el contrario, *Ericsson y Simon* mencionan algunas ventajas que poseen estos reportes orales, como por ejemplo, que los participantes no necesitan ser entrenados antes de realizar las actividades que se les indican y que sus reportes verbales son consistentes con la estructura de los procesos cognitivos habituales.³⁶

Se concluye que existen diversas metodologías que permiten determinar la validez de contenido, dentro de ellas se destaca la evaluación por comité de expertos y el pensamiento hablado o entrevista cognitiva.

En la primera de ellas, es fundamental la correcta selección de ellos (características y número), la utilización adecuada de la escala Likert y su evaluación de acuerdo a estándares establecidos. Así también, existe la opción de seleccionar alguna técnica que guíe de manera más concreta el proceso de validación según el objetivo de ella, como por ejemplo Metodología Q, Modelo de Fehring y Método Delphi.

La entrevista cognitiva es una técnica que requiere que los participantes piensen en voz alta mientras llevan a cabo la actividad solicitada; el relato que se obtiene se graba para ser posteriormente transcrito y analizado. Se deben tener presente los altos costos que requiere esta técnica.

Para un mejor resultado de validez de contenido se sugiere utilizar más de una metodología, con el fin de complementarlas, aumentando así la rigurosidad del proceso.

REFERENCIAS BIBLIOGRÁFICAS

1. Galan Perroca M. Desarrollo y validación de contenido de la nueva versión de un instrumento para clasificación de pacientes. *Rev Latino-Am Enfermagem*. 2011;19(1):1-9.
2. López A, Ropero J, Peralta JC. Estudio de validez del examen de Estado Saber 11 de inglés. *FOLIOS*. 2011;34:77-91.
3. Leyva J. Una reseña sobre la validez de constructo de pruebas referidas a criterio. *Perfiles educativos*. 2011;33(131):131-54.
4. Piratoba BN. Confiabilidad del instrumento para medir la "habilidad de cuidado de cuidadores familiares de personas con enfermedad crónica." en *Cuidadores de personas mayores de la localidad de Usaquen, Bogota, D.C [tesis magíster]*. Bogotá: Universidad Nacional de Colombia, Facultad de Enfermería; 2011.
5. Gómez J, Hidalgo M. La validez en los tests, escalas y cuestionarios. Universidad de Antioquia. Facultad de Ciencias Sociales y Humanas. Centro de Estudios de Opinión. S.f. [citado 2 Ene 2013] Disponible en: <http://aprendeenlinea.udea.edu.co/revistas/index.php/ceo/article/viewFile/1750/1370>
6. Delgado-Rico E, Carretero-Dios H, Ruch W. Content validity evidences in test development: An applied perspective. *Intern J Clin Health Psych*. 2012;12(3):449-60.
7. Prieto G, Delgado A. Fiabilidad y validez. *Papeles del psicólogo*. 2012;31(1):67-74.
8. Padilla JL, Gómez J, Hidalgo MD, Muñiz J. La evaluación de las consecuencias del uso de los tests en la teoría de la validez. *Psicothema*. 2006;18(2):307-12.
9. Perpiñá-Galvañ J, Richart-Martínez M, Cabañero-Martínez MJ, Martínez-Durá I. Validez de contenido de versión corta de la subescala del Cuestionario State-Trait Anxiety Inventory (STAI). *Rev Latino-Am Enfermagem*. 2011;19(4):[06 pantallas]. [citado 2 Ene 2013] Disponible en: http://rua.ua.es/dspace/bitstream/10045/24995/2/2011_Perpina_etal_RLAE_esp.pdf

10. Carvajal A, Centeno C, Watson R, Martínez M, Rubiales A. ¿Cómo validar un instrumento de medida de la salud? *An Sist Sanit Navar*. 2011;34(1):63-72.
11. Abad F, Garrido J, Olea J, Ponsada 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.
12. Barrazas A. La consulta a expertos como estrategia para la recolección de evidencias de validez basadas en contenido. *Investigación Educativa Duranguense*. 2007;7:5-13.
13. Rodríguez Martínez C, Sossa MP. Validación de un cuestionario de conocimientos acerca de asma. *Revista Colombiana de Neumología*. 2004;16(3):162-8.
14. García S. La validez y la confiabilidad en la evaluación del aprendizaje desde la perspectiva hermenéutica validity and reliability in the evaluation of learning from a hermeneutical perspective. *Rev Ped*. May. 2002;23(67):297-318.
15. García López E, Cabero Almenara J. Diseño y validación de un cuestionario dirigido a describir la evaluación en procesos de educación a distancia. *Educec-e*. 2011;35:1-26.
16. Backhoff Escudero E, Aguilar Villalobos J, Larrazolo Reyna N. Metodología para la validación de contenidos de exámenes normativos. *Rev Mex Psicol*. 2006;23(1):79-86.
17. Yaghmale F. Content validity and its estimation. *J Med Educ*. 2003;3(1):25-7.
18. Cisneros E, Jorquera M, Aguilar A. Validación de instrumentos de evaluación docente en el contexto de una universidad española. *Voces y Silencios: Revista Latinoamericana de Educación*. 2012;3(1):41-55.
19. Messick S. Validity. En: Linn RL, editors. *Educational measurement*. New York: American Council on Education and Macmillan Educational measurement; 1989. p. 13-103.
20. Sánchez R, Navarro A, Rueda-Jaimes GE, Gómez-Restrepo C. Desarrollo y validación de la versión II de la escala EMUN. *Rev Colomb Psiquiat*. 2011;40(4):647-59.
21. Gajewski BJ, Price LR, Coffland V, Boyle DK, Bott MJ. Integrated analysis of content and construct validity of psychometric instruments. *Qual Quant*. 2013;47:57-78.
22. 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;31(1):59-70.
23. Chacon S, Perez-Gil JA, Holgado Tello FP, Lara Ruiz A. Evaluación de la calidad universitaria: validez de contenido. *Psicothema*. 2001;13(2):294-301.

24. García Ñ, Fernandez SJ. Procedimiento de aplicación de trabajo creativo en grupo de expertos. *Energética*. 2008;29(2):46-50.
25. 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(2):90-5.
26. Ortega Mohedano F. El método Delphi, prospectiva en ciencias Sociales. *Revista Ean*. 2008;65:31-64.
27. Lepos Ferrari A, Campos P, Andres Felli VE, Coggon D. Traducción, adaptación y validación del cuestionario Cultural and Psychosocial Influences on Disability (CUPID) para uso en Brasil. *Rev Latino-Am Enfermagem*. 2010;18(6):1-7.
28. Nuviola A, Tamayo JA, Iranzo J, Falcón D. Creación, diseño, validación y puesta en práctica de un instrumento de medición de la satisfacción de usuarios de organizaciones que prestan servicios deportivos. *Retos: Nuevas tendencias de educación física, deporte y recreación*. 2008;14:10-6.
29. Bonilla CP. Diseño de un instrumento para evaluar los factores que influyen en la adherencia a tratamiento, en personas que presentan los factores de riesgo de enfermedad cardiovascular. *Av Enferm*. 2007;25(1):46-55.
30. Guirao-Goris JA. Investigación en nomenclatura enfermera. Cap. 3. En: Guirao-Goris JA, Camaño RE, Cuesta A. Diagnóstico enfermero. Categorías, formulación e investigación [en Internet]. Sevilla: Ediversitas Multimedia; 2001. p. 1-24. [citado 2 Ene 2013] Disponible en: http://www.uv.es/joguigo/materiales-dominio-profesional/investigacion_dde.pdf
31. Reis Dos Santos S, Schor N. Vivências da maternidade ena adolescencia precoce. *RevSaúde Pública*. 2003;37(1):15-23.
32. Almansa C, Rey E, Bolaños E, Palma M, Álvarez A, Díaz-Rubio M. Opinión de los médicos españoles sobre el síndrome de intestino irritable: resultados de un estudio utilizando el método Delphi. *Rev Esp Enferm Dig*. 2007;99(4):210-7.
33. Calaboug F, Crespo J. Uso del método Delphi para la elaboración de una medida de la calidad percibida de los espectadores de eventos deportivos. *Retos: Nuevas tendencias de educación física, deporte y recreación*. 2007;15:21-5.
34. Gil-Gomez B, Pascual-Ezama D. La metodología Delphi como técnica de estudio de la validez de contenido. *Anpsicol*. 2012;28(3):1011-20.
35. Armengol L. Los protocolos de pensamiento en voz alta como instrumento para analizar el proceso de escritura. *RESLA*. 2007;20:27-35.
36. Ericsson A, Simon H. Protocol analysis. Cambridge: MIT Press, 1993. p. 1-62.
37. Buteler L, Gangoso Z. Algunos aspectos metodológicos de la investigación en resolución de problemas en física: una revisión. *Ciência & Educação (Bauru)*. 2008;14(1):1-14.

38. Hannu K, Pallab P. A comparison of concurrent and retrospective verbal protocol analysis. *Am J Psych.* 2000; 113(3):387-404.
39. Torrealba C, Rosales L. El protocolo oral como vía para la indagación del conocimiento metacognitivo: análisis De experiencias de investigación. *Investigación y Postgrado.* 2008;28(1):10-16.
40. Villegas J, Castro E, Gutiérrez J. Representaciones en resolución de problemas: Un estudio de caso con problemas de optimización. *Electr J Res Educ Psych.* 2009; 17(1): 279-308.
41. De Brito DM. La verbalización como registro para análisis en la investigación sobre lectura. *Anales de Documentacion.* 2006; (9): 43-51.

Recibido: 7 de noviembre de 2013.

Aprobado: 5 de enero de 2014.

MSc. *Marcela Urrutia Egaña*. Dirección: Pontificia Universidad Católica de Chile. Avenida Vicuña Mackenna No. 4860. San Joaquín. Santiago de Chile. Dirección postal: 782-0436. marcelaurrutia@uc.cl