

Validez y confiabilidad del cuestionario "Prácticas de cuidado que realizan consigo mismas las mujeres en el posparto"

Validity and reliability of the questionnaire "Self-care practices of women during postpartum period"

Validez e confiabilidade do questionário "Práticas de autocuidado das mulheres durante o pós-parto"

CAROLINA VARGAS PORRAS*, LUZ MERY HERNÁNDEZ MOLINA**

Resumen

Todo instrumento de recolección de datos debe ser válido y confiable para que enfermería intervenga efectivamente y le permita proponer políticas de salud en búsqueda de reducir las cifras de morbilidad materna y perinatal.

El objetivo general de la investigación fue determinar la validez y confiabilidad del cuestionario "Prácticas de cuidado que realizan consigo mismas las mujeres en el posparto". Es una investigación metodológica. La validez de contenido mediante la prueba de expertos reportó un puntaje de 0,88, la validez facial con 10 expertas y 10 mujeres en el posparto evidenció consenso de acuerdos en relación con la claridad y comprensión del cuestionario. Se realizó prueba piloto para determinar posibles ajustes. Participaron 176 mujeres lactantes en el posparto de diferentes estratos socioeconómicos, residentes en

Bogotá, a quienes se les aplicó, previo consentimiento informado, un cuestionario conformado por 32 ítems.

Se elaboró una base de datos en Excel y los datos se procesaron en SPSS versión 12.0. La validez de escala evidenció que es una escala corta y la validez discriminante determinó 12 ítems que tenían mayor correlación y agrupó los que tenían correlaciones bajas. La confiabilidad del cuestionario de 32 ítems presentó alfa de Cronbach de 0,59 y el cuestionario de 12 ítems alfa de Cronbach de 0,68, considerándose un puntaje aceptable y bueno en los primeros estudios de validación de un instrumento. La confiabilidad del cuestionario permea todos los procesos de validación, por lo que su interpretación debe ir de la mano del perfeccionamiento continuo de la validez.

Palabras clave: confiabilidad y validez, cuestionario, atención de Enfermería, mujeres, periodo de posparto (Fuente: DeCS, BIREME).

Abstract

All data collection instruments must be valid and reliable for the nurse to be able to intervene effectively enabling her to propose health policies that pursue the reduction of maternal and perinatal morbidity and mortality.

The general objective of this study was to determine the validity and reliability of the questionnaire named "Self-care practices of women during postpartum period". It is a methodological research. Content validity conducted through the experts test reported a score of 0,88; the face validity with 10

* Enfermera Universidad Industrial de Santander, magíster en Enfermería con énfasis en Cuidado para la Salud Materno Perinatal, Universidad Nacional de Colombia, profesora de la Universidad Industrial de Santander, cvargasporras@yahoo.com, Bucaramanga, Colombia.

** Docente Universidad Nacional de Colombia, magíster en Enfermería con énfasis en Cuidado para la Salud Materno Perinatal. Directora de tesis, lmhernandezmo@unal.edu.co, Bogotá, Colombia.

experts and 10 women in a postpartum stage evidenced a consensus of agreement in relation with the clarity and comprehension of the questionnaire. A pilot test was conducted to determine possible adjustments. With participation of 176 postpartum nursing mothers from different socio-economic strata, residing in Bogotá, to whom, prior informed consent, a 32 item questionnaire was applied.

A database was put together in Excel and data were processed in SPSS version 12.0. The scale validity showed that it is a short scale; discriminating validity determined 12 items that had a greater correlation and grouped those that had low correlations. The reliability of the 32 item questionnaire presented a Cronbach's alpha of 0,59 and the 12 item questionnaire a Cronbach's alpha of 0,68, which is considered an acceptable and a good score in the first validation studies of an instrument. The reliability of the questionnaire permeates all validation processes, reason why its interpretation must go hand in hand with continued validity improvement.

Key words: reproducibility of results, questionnaire, nursing care, women, postpartum period

Resumo

Qualquer instrumento de coleta de dados deve ser válido e confiável no intuito de garantir que a intervenção de enfermagem seja efetiva e possibilite a proposição de políticas de saúde em busca de reduzir as cifras de morbimortalidade materna e perinatal.

O objetivo geral da pesquisa foi determinar a validade e confiabilidade do questionário "Práticas de autocuidado das mulheres durante o pós-parto". É uma pesquisa metodológica. A validade do conteúdo mediante a prova de especialistas apresentou uma pontuação de 0,88, a validade facial com 10 especialistas e 10 mulheres no pós-parto evidenciou consenso no tocante à clareza e compreensão do questionário. Realizou-se uma prova piloto para determinar possíveis ajustes. Participaram 176 mulheres lactantes no pós-parto de diferentes níveis socioeconômicos, residentes em Bogotá, que, com prévia autorização, completaram o questionário formado por 32 itens.

Elaborou-se um banco de dados em Excel e os dados foram processados em SPSS versão 12.0. A validade de escala evidenciou que é uma escala curta e a validade discriminante determinou os 12 itens com maior correlação e agrupou todos os itens que tinham correlações baixas. A confiabilidade do questionário de 32 itens apresentou um alfa de Cronbach de 0,59 e o questionário de 12 itens alfa de Cronbach de 0,68. A pontuação se considera aceitável e boa nos primeiros levantamentos de validação de um instrumento. A confiabilidade do questionário passa por todos os processos de validação, é por isso que a interpretação deve ir acompanhada do aprimoramento contínuo da validade.

Palabras clave: reprodutibilidade dos testes, questionários, cuidados de enfermagem, mulheres, período pós-parto

INTRODUCCIÓN

Los instrumentos que se utilizan en el proceso de investigación son herramientas fundamentales que proporcionan la información; por consiguiente, todo instrumento de recolección de datos debe ser válido y confiable.

Al no existir instrumentos válidos y confiables que permitan describir las prácticas de cuidado que realizan consigo mismas las mujeres en el posparto, Lucy Muñoz de Rodríguez y Carmen Helena Ruiz de Cárdenas, profesoras de la Facultad de Enfermería de la Universidad Nacional de Colombia, elaboraron un cuestionario que responde a esta necesidad y que mide seis dimensiones: actividad/ejercicio y descanso, alimentación, higiene y cuidados personales, hábitos no beneficiosos, medicamentos y sistemas de apoyo; su validez y confiabilidad se hacen indispensables para poder aplicarlo porque, en la medida en que se utilicen instrumentos válidos y confiables, se fortalece el cuerpo de conocimientos de la profesión, a través de procesos investigativos que sustentan científicamente el ser y el quehacer de enfermería.

La presente investigación corresponde a un estudio metodológico a través del cual se determinó la validez de contenido, validez facial, validez discriminante, validez de escala y la confiabilidad del cuestionario "Prácticas de cuidado que realizan consigo mismas las mujeres en el posparto", de 32 ítems, que utiliza una escala tipo Likert, dando como resultado una versión mejorada de 12 ítems y constituyéndose en un aporte muy valioso, desde el rigor psicométrico, para la línea de investigación "Salud sexual y reproductiva desde la perspectiva transcultural", del grupo académico Cuidado de Enfermería Materno Perinatal.

PROBLEMA

¿Cuál es la validez y confiabilidad del cuestionario "Prácticas de cuidado que realizan consigo mismas las mujeres en el posparto"?

OBJETIVO GENERAL

Determinar la validez y confiabilidad del cuestionario "Prácticas de cuidado que realizan consigo mismas las mujeres en el posparto".

OBJETIVOS ESPECÍFICOS

- Determinar la validez de contenido del cuestionario mencionado.
 - Determinar la validez facial del cuestionario.
 - Determinar la confiabilidad del cuestionario.

DEFINICIÓN DE CONCEPTOS

Validez de un instrumento. Grado en que un instrumento mide la variable que pretende medir (1).

Validez de contenido. Grado en que un instrumento refleja el dominio específico del contenido de lo que se mide (1).

Validez aparente o facial. Grado en que aparentemente un instrumento mide la variable en cuestión; se encuentra vinculada con la validez de contenido (1).

Como su nombre lo indica, esta evaluación se interesa por los aspectos formales de la medición (el instrumento y el procedimiento) y los juzga desde el conocimiento vigente. Una medición se considera aparentemente válida si, a la luz de lo que se conoce sobre el tema, sus resultados pudieran reflejar los atributos del fenómeno en estudio. Este juicio se obtiene mediante la evaluación del instrumento por un grupo de jueces en general expertos, quienes, con base en su experiencia y en las teorías aceptadas sobre el tema, conceptúan si la medición podrá reflejar de manera plausible los atributos en cuestión (2).

Confiabilidad de un instrumento. Grado de congruencia con que un instrumento mide el atributo para el que está diseñado. A menor variación de las mediciones repetidas de un instrumento, mayor la confiabilidad (1).

Prácticas de cuidado que realizan consigo mismas las mujeres en el posparto. Actividades que realizan las mujeres con ellas mismas durante el posparto, con el fin de seguir viviendo, mantener la salud, atender sus enfermedades y conservar su bienestar; basadas en un sistema de creencias y tradiciones culturales, difícilmente remplazadas por nuevos conocimientos y además desconocidas por quienes ofrecen cuidado (3).

METODOLOGÍA

Es un estudio metodológico con abordaje cuantitativo. La muestra estuvo conformada por mujeres lactantes

adultas en el posparto, residentes en Bogotá, de los diferentes estratos socioeconómicos, cuyo parto fue atendido en los hospitales de Engativá, La Victoria y usuarias de Unisalud. El tamaño de la muestra se determinó después de realizar la validez facial y de contenido, teniendo en cuenta la medición de las propiedades psicométricas del cuestionario; dio un total de 176 mujeres en el posparto, número representativo que se encuentra dentro del rango necesario establecido para el cálculo de la confiabilidad mediante el coeficiente alfa de Cronbach (4-6) y además definido como cinco participantes por el número de ítems del cuestionario validado ($5 \times 32 = 160$). Cinco (7, 8) es el mínimo de sujetos por ítem recomendado para establecer la validez factorial de un cuestionario (9-11), más el 10% para reponer posible atrición y que en una futura investigación se pueda realizar con esta base de datos la validez de constructo del cuestionario. Teniendo en cuenta la estratificación de la población en Bogotá, se calcularon los porcentajes correspondientes de la muestra total para que fueran representativos.

Criterios de inclusión: mujeres lactantes mayores de edad que se encuentren entre 1 y 180 días después del parto, residentes en Bogotá, y mujeres primíparas que tienen por lo menos 8 días posparto. **Criterios de exclusión:** mujeres con óbito fetal o mortinato, mujeres con alteraciones mentales, mujeres con alteraciones de la comunicación, mujeres hospitalizadas por enfermedad durante el puerperio, mujeres que presenten síntomas de depresión posparto.

De los aspectos éticos se tuvieron en cuenta los siguientes: el aval del comité de ética de la Facultad de Enfermería, las pautas éticas internacionales 4, 5, 6 y 8, para la investigación biomédica en seres humanos recomendadas por CIOMS, los lineamientos de la Resolución 008430, artículos 1 a 37, los aspectos relacionados con investigación de la Ley 911, capítulo IV del Título III, artículos 29-30, las recomendaciones éticas de la Facultad de Enfermería. Se clasifica como una investigación sin riesgo y se realizó consentimiento informado de manera verbal y escrita. Los principios éticos de la investigación fueron: fidelidad, justicia, veracidad, reciprocidad, respeto a la autonomía, beneficencia y no maleficencia.

Procedimiento

Fase I. Revisión exhaustiva de la literatura científica, adaptación del cuestionario base y descripción del cuestionario inicial.

Fase II. Determinación de la validez de contenido del cuestionario, determinación de la validez facial del cuestionario, cuestionario modificado, prueba piloto.

Fase III. Aplicación del cuestionario modificado, procesamiento y análisis de los datos, determinación de la confiabilidad del cuestionario, cuestionario final y presentación de los resultados.

RESULTADOS Y ANÁLISIS DE LA VALIDEZ DE CONTENIDO

La validez de contenido se determinó a través de la prueba de expertos descrita por Polit y Hungler (12), en la que participaron 10 expertas en el área materno perinatal, quienes evaluaron cada uno de los ítems del cuestionario teniendo en cuenta dos criterios: *pertinencia*, es decir, el ítem evalúa lo que se pretende evaluar, y *relevancia*, definida como la importancia del ítem para la evaluación de las prácticas de cuidado; se tuvo en cuenta la siguiente escala:

Pertinencia: 1. no pertinente, 2. poco pertinente, 3. pertinente, 4. muy pertinente.

Relevancia: 1. no relevante, 2. poco relevante, 3. relevante, 4. muy relevante.

Análisis cuantitativo de validez de contenido

- Se calculó el *índice de validez de contenido (IVC) para cada experta*, dado por la siguiente fórmula:

$$\frac{\text{Número de ítems con puntuación entre 3 y 4}}{\text{Número total de ítems}}$$

Los valores de índice de validez de contenido para cada experta demostraron, en su gran mayoría, validez de contenido elevada, la cual, según Polit y Hungler (12), incluye los puntajes iguales o mayores de 0,80; incluso una de ellas evaluó con el puntaje máximo de validez del cuestionario. En este índice de validez solo dos expertas calificaron por debajo de este valor (0,58 y 0,76), lo que se contrasta posteriormente con la evaluación cualitativa que emitieron al respecto.

- El índice de *validez de contenido general* se calculó según la siguiente fórmula:

$$\frac{\text{Suma del IVC calculado para cada una de las expertas}}{\text{No. de expertas}}$$

El índice de validez de contenido general (IVCG) mayor o igual a 0,80, según Polit y Hungler (12), se considera indicativo de validez de contenido elevada, como ocurre en este caso cuyo puntaje fue de 0,88, mayor que el valor mínimo exigido, de 0,62 según Lawshe (13) para la prueba con 10 expertos, lo cual refleja que los ítems del cuestionario miden un dominio específico fundamentado desde la literatura científica relacionada con las prácticas de cuidado de las mujeres en el posparto, garantizando la validez de contenido general.

- El índice de *validez de contenido de cada ítem* se determinó con el fin de evaluar la pertinencia y relevancia de cada uno de ellos. El cálculo se realizó mediante la siguiente fórmula:

$$\frac{\text{Número de expertas que concuerdan en el valor de relevancia o pertinencia en cada ítem}}{\text{Número total de expertas}}$$

Teniendo en cuenta que la pertinencia consiste en que el ítem evalúe lo que se pretende evaluar, se encontró que los ítems 1, 6, 7, 11, 19, 28, 29 y 30 obtuvieron puntajes menores de 0,80, por lo cual se hace necesario contrastarlos con la evaluación cualitativa de las expertas. Por otro lado, se resalta que los ítems 12, 13, 15, 22, 23, 24, 25, 27, 31, 32 y 33 fueron calificados con el valor máximo de pertinencia y que los restantes ítems del cuestionario puntuaron con validez de pertinencia elevada.

Teniendo en cuenta que la relevancia se define como la importancia del ítem para la evaluación de las prácticas de cuidado, se encontró que los ítems 1, 19 y 30 obtuvieron un valor inferior a 0,80, por lo que se hace necesario contrastarlos con la evaluación cualitativa de las expertas. Por otro lado, se resalta que los ítems 4, 6, 12, 13, 15, 16, 22, 23, 24, 25, 27, 31, 32 y 33 fueron calificados con el valor máximo de relevancia y que los restantes ítems del cuestionario puntuaron con validez de relevancia elevada.

Los resultados de la validez de contenido de cada dimensión presentaron cinco dimensiones con validez de contenido elevada y una con validez de contenido buena.

En general, se puede concluir que el análisis cuantitativo del cuestionario pone en evidencia que este presenta validez de contenido tanto en pertinencia como en relevancia.

Tabla 1. Validez de contenido por cada dimensión.

Dimensión	Validez de contenido
Actividad / Ejercicio y descanso	0,85
Alimentación	0,84
Higiene y cuidados personales	0,90
Hábitos no beneficiosos	0,97
Medicamentos	0,77
Sistemas de apoyo	0,96

Fuente: dimensiones de cuidado del cuestionario "Prácticas de cuidado que realizan consigo mismas las mujeres en el posparto".

Análisis cualitativo de la validez de contenido y de la validez facial

Se efectuó teniendo en cuenta los siguientes parámetros:

Para realizar modificaciones en el ítem, se deben cumplir los siguientes parámetros:

- La modificación debe facilitar la claridad y comprensión del ítem.
- Debe existir un consenso de acuerdos entre las expertas igual o mayor de 50%.

Para eliminar un ítem, se deben cumplir los siguientes parámetros:

- Debe existir un consenso de acuerdos entre las expertas igual o mayor de 60%.
- El ítem debe haber obtenido un puntaje menor de 0,80 según el promedio de pertinencia y relevancia.
- No existe evidencia científica que fundamente su consideración dentro del cuestionario, para medir prácticas de cuidado durante el posparto.

RESULTADOS Y ANÁLISIS DE LA VALIDEZ FACIAL

Se determinó teniendo en cuenta la evaluación cualitativa realizada según los criterios de claridad y

comprensión. En esta validez participaron las mismas 10 expertas que hicieron la validez de contenido, las cuales evaluaron cualitativamente cada ítem y el cuestionario en general según la claridad y la comprensión, y 10 mujeres en el posparto, que evaluaron cualitativamente la claridad y comprensión del cuestionario en general.

Teniendo en cuenta lo anterior, las expertas y mujeres en el posparto manifestaron la importancia, claridad y comprensión de los ítems y del cuestionario en general. 80% de expertas expresaron claridad y comprensión del cuestionario en general y 90% de mujeres en el posparto expresaron claridad y comprensión del cuestionario en general, para un total de 85% de acuerdos sobre claridad y comprensión, que se considera un alto porcentaje de acuerdos para la validez facial.

RESULTADOS Y ANÁLISIS DE LA VALIDEZ DE ESCALA

La validez de escala se determinó según Tristán (14), que evalúa la calidad de una prueba en la cual el foco principal de los diseños centrados en la validez es el conjunto formal de objetivos que son característicos de un dominio teórico (15).

Una vez efectuada en Excel la base de datos de las 176 mujeres en posparto, quienes cumplieron con los criterios de inclusión de la investigación, se realizó el análisis de la escala del cuestionario, mediante la graficación e interpretación de las medias de los ítems del total del cuestionario. La validez de escala se utilizó para focalizar en la calidad del criterio teórico que sustenta la escala y conocer objetivamente si la distribución de los ítems de cada una de las dimensiones teóricas de la prueba corresponde con un criterio teórico externo.

Para revisar las cualidades de la validez de escala descritas, se dispuso de una herramienta gráfica en Excel, que consistió en representar en el plano las medias ascendentes de los ítems (media de respuestas positivas en cada ítem) versus los ítems organizados según media ascendente (los ítems totales del cuestionario, ordenándolos de menos a más). Si se cuenta con una escala perfecta, todos los ítems deberían quedar alineados en una recta que va de $p_{inf} = 0,2$ a $p_{sup} = 0,8$ (valores teóricos sugeridos 20%-80%).

Se compararon las posiciones reales de los ítems con las posiciones teóricas y se obtuvo la diferencia absoluta

media, la cual no debía ser mayor de 10% (DAM = 0,10) y que se define por:

$$DAM = \frac{\sum_{i=1}^M p_i - p_e}{M}$$

donde:

M = número de ítems

p_i = media de respuestas positivas en el ítem i

p_e = $p \text{ inf} + (p \text{ sup} - p \text{ inf}) \frac{i-1}{M-1}$, media esperada

en la recta que va de $p \text{ inf}$ a $p \text{ sup}$.

Cuando los ítems se alejaban significativamente de una distribución rectilínea, se puso en tela de juicio si se trataba de una variable que define una escala.



Figura 1. Validez de escala ajustada a 0,8-2,3.

Fuente: ítems y medias de los ítems del cuestionario "Prácticas de cuidado que realizan consigo mismas las mujeres en el posparto".

En Excel se ordenaron en forma ascendente las medias de cada uno de los ítems del cuestionario y con una diferencia absoluta media de 0,12, el mejor ajuste fue de 0,8 a 2,3, lo que quiere decir que es una escala corta comparada con la recta teórica 20-80, que permite la distribución de las dificultades de los ítems; por tanto, el cuestionario necesita completar o mejorar los ítems, para medir el rango más amplio posible de las prácticas de cuidado que realizan consigo mismas las mujeres en el posparto.

La escala se encuentra centrada en 0,015 (promedio entre el límite inferior y el superior) del atributo medido que, si se compara con el valor medio perfecto en 0,5, se puede concluir que el rasgo se presenta en los sujetos en 48% menos de lo esperado. Vale la pena aclarar que este hallazgo no es posible interpretarlo desde la

direccionalidad con que fue medido el fenómeno ya que el sentido métrico es diferente por dimensión teórica abordada y la validez de escala no se presenta por dimensión sino por prueba total debido a su poca utilidad por longitud de prueba.

Es por ello que el anterior resultado se utiliza con fines estadísticos, dando evidencia que se trata de una escala que se queda corta al no poseer ítems con más del 2,3% de respuestas positivas.

Para obtener la correlación entre la recta teórica y la recta empírica se utilizó básicamente la fórmula producto-momento de Pearson, dada por:

$$r = \frac{\sum xy}{N\sigma_x\sigma_y}$$

donde:

r = correlación entre las variables X , Y

x = desviación de una puntuación X respecto a la media $M_x = (X - M_x)$

y = desviación de una puntuación Y respecto a la media $M_y = (Y - M_y)$

σ_x = desviación estándar de la distribución de puntuaciones X

σ_y = desviación estándar de la distribución de puntuaciones Y

En la figura 1 se observa que la recta empírica no se ajusta a la recta teórica 20-80; asumiendo una DAM de 0,12 para un ajuste 0,8-2,3, se obtiene una correlación de Pearson de 0,94, lo cual permite certificar que existe una correlación casi perfecta entre los datos y la recta teórica sugerida. Pese a la correlación obtenida y considerando las falencias métricas de escala, se puede afirmar que queda en tela de juicio que el cuestionario diseñado sea una escala verdadera. Lo anterior debe someterse a verificación en futuros estudios.

RESULTADOS Y ANÁLISIS DE LA VALIDEZ DISCRIMINANTE

Esta validez se enfocó en identificar la variable que mide el cuestionario, en contraposición con las otras variables que pudieran estar presentes, pero que corresponden a otros rasgos o constructos. En este estudio se decidió utilizar el análisis de la correlación ítem-total (correlación punto-biserial) (14), definido por:

$$rpbis = \frac{Mp - Mq}{St} \sqrt{pq}; 0 \leq rpbis \leq 1$$

donde:

- $rpbis$ = Correlación punto-biserial, igual a la correlación producto-momento de Pearson.
- Mp = Media de puntuaciones de los sujetos que contestan correctamente al ítem.
- Mq = Media de puntuaciones de los sujetos que contestan incorrectamente al ítem.
- St = Desviación estándar de las puntuaciones de los sujetos.
- p = proporción de respuestas correctas en el ítem, igual a la proporción de casos que están en el grupo superior en el ítem.
- q = proporción de respuestas incorrectas en el ítem ($q = 1-p$), igual a la proporción de casos que están en el grupo inferior en el ítem.
- pq = varianza del ítem.

El análisis de la correlación ítem-total (correlación punto-biserial) se realizó según los siguientes tres casos posibles:

- Correlación positiva (superior a 0,196)
- Correlación negativa (inferior a 0,196 y negativos)
- Correlación baja (inferior a 0,196)

Se consideró que cada una de estas agrupaciones discriminaba si cumplía con los siguientes criterios de los conjuntos de ítems (agrupados en cualquiera de estos tres casos); se considera como una posible variable si se satisface que:

- La variable cuenta con cuatro ítems por lo menos.
- Los ítems correlacionan positivamente con el conjunto (valores superiores a 0,196).
- El conjunto produce valores altos del coeficiente de consistencia interna, alfa de Cronbach con un valor mínimo aceptable según la teoría clásica de 0,60 (16).

La correlación ítem-total es de gran relevancia porque indica la correlación lineal entre el ítem y el puntaje total. Teniendo en cuenta la correlación ítem-total que reportaba cada ítem del cuestionario mediante el programa estadístico SPSS versión 12.0, se establecieron tres grupos de discriminación: uno conformado por las correlaciones negativas, otro por las correlaciones menores de 0,196 y el último por las correlaciones mayores de 0,196.

Se consideraron correlaciones de ítem-total superiores a 0,196 (14), valor aceptado para estudios en el

área de la salud como un indicador de la pertinencia de un ítem al cuestionario; por lo tanto, es más un índice de validez del ítem dentro de una prueba.

Las correlaciones mayores de 0,196 presentaron un alfa de Cronbach de 0,68 y las correlaciones ítem-total fueron altas entre sí, evidenciando que estos ítems al presentar mejor consistencia interna aumentan la confiabilidad.

RESULTADOS Y ANÁLISIS DE LA CONFIABILIDAD

Como medida de la calidad de una prueba, se exige que tenga confiabilidad, la cual se refiere a que en mediciones repetidas se tengan resultados similares o que la medición se realice con la mayor precisión posible (14).

La consistencia interna busca determinar si las respuestas de los sujetos son consistentes a lo largo de la prueba y el coeficiente alfa de Cronbach se enfoca en la consistencia interna de las respuestas de los sujetos (14). El coeficiente alfa de Cronbach se define por:

$$\alpha = \frac{N}{N-1} \left(1 - \frac{\sum Vi}{Vk} \right)$$

donde:

- N = número de ítems
- Vi = varianza del ítem i
- Vk = varianza de los puntajes brutos de los sujetos

El coeficiente alfa se enfoca en la consistencia interna de las respuestas de los sujetos. Según Tristán (14), sus valores fluctúan entre [-8, 1], no entre [0, 1] que es lo correcto por ser un cociente de varianzas. Alfa tiende a -8 si $N = 1$, en cambio alfa tiende a 1 cuando N tiende a $+\infty$, independientemente de la calidad de los ítems, pero también cuando los ítems son idénticos o muy homogéneos, independientemente del número de ítems (14). En contraste, Hernández et ál. (17) consideran que el coeficiente de confiabilidad oscila entre [0, 1] (17).

Existe una gran discusión entre diferentes autores sobre el valor mínimo considerado aceptable, bueno y alto para el coeficiente de confiabilidad alfa de Cronbach.

Tristán (14) sugiere que la confiabilidad no tiene que ser alta sino satisfacer primero la validez y la objetividad, para después contar con la confiabilidad pertinente.

Por otro lado, Ruiz (18) propone una manera para establecer los rangos de magnitud del coeficiente, que son: 0,81 a 1,00, muy alta; 0,61 a 0,80, alta; 0,41 a 0,60, moderada; 0,21 a 0,40, baja y 0,01 a 0,20, muy baja. A diferencia de Hernández et ál. (17) que sugieren que 0,25 indica una baja confiabilidad, 0,50 es media o regular, si supera el 0,75 es aceptable y si es mayor de 0,90 es elevada.

Polit y Hungler (12), al igual que Burns y Grove (19), afirman que no hay normas para determinar qué coeficiente de confiabilidad resulta aceptable, pero que en general es aceptable hasta un valor mínimo de 0,70.

Otros autores (20-26) consideran un coeficiente de confiabilidad de alfa de Cronbach aceptable mínimo de 0,6.

La confiabilidad del cuestionario se calculó mediante el coeficiente alfa de Cronbach utilizando el programa estadístico SPSS versión 12.0; se obtuvo el siguiente resultado:

Tabla 2. Alfa de Cronbach, cuestionario de 32 ítems.

Alfa de Cronbach	No. de ítems
0,59	32

Fuente: confiabilidad del cuestionario "Prácticas de cuidado que realizan consigo mismas las mujeres en el posparto".

Por tanto, para la interpretación de este puntaje se tuvo en cuenta que el cuestionario mide 6 dimensiones y que según Ruiz (18) las pruebas multidimensionales, es decir, aquellas que miden diferentes aspectos o dimensiones de un constructo, están formadas por subpruebas. Este tipo de instrumentos tienden a presentar coeficientes de confiabilidad de consistencia interna de magnitud moderada, indicando que no son totalmente homogéneas; sin embargo, sus respectivas subpruebas pueden serlo, lo cual se evidenció igualmente en una investigación (27) en que el coeficiente alfa de Cronbach fue de 0,66 y cuyo cuestionario estaba también constituido por 6 dimensiones.

PROPUESTA DE MEJORAMIENTO: CUESTIONARIO FINAL

Según Nunnally y Berstein (28), en las primeras etapas de la investigación de la validación puede ahorrarse

tiempo y energía mediante instrumentos que solo tienen una confiabilidad modesta. Posteriormente se aplican correcciones y, si estas sugieren buenos resultados, valdrán la pena el tiempo y el esfuerzo para incrementar el número de reactivos y tratar de reducir el error de medición. Por lo tanto, en esta investigación se propone un cuestionario de 12 ítems, los cuales presentaron un alto poder de discriminación y consistencia interna y cuyo coeficiente alfa de Cronbach es mayor que la confiabilidad reportada por el cuestionario inicial de 32 ítems.

Tabla 3. Comparación de cuestionarios.

Confiabilidad	Cuestionario con 32 ítems	Cuestionario con 12 ítems
Alfa de Cronbach	0,59	0,68

Fuente: confiabilidad del cuestionario "Prácticas de cuidado que realizan consigo mismas las mujeres en el posparto" con 32 ítems y con 12 ítems.

Se observó en este caso que al retirar reactivos con bajas correlaciones (como se ha evidenciado en otras investigaciones) (29-31) del cuestionario inicial de 32 ítems, aumentó la confiabilidad, dando como resultado un cuestionario de 12 ítems.

Error estándar de medida (14)

Conocida la confiabilidad, se deduce una expresión para determinar el error de medida del instrumento, parámetro necesario para establecer la calidad de la prueba.

La prueba y sus componentes deben tener un error estándar de diseño no mayor de 5% o, por revisión, de 10%. En caso de superar este valor, deberá justificarse en función de las cualidades de la prueba, sus dimensiones y el impacto en los sustentantes (14), pero en este cuestionario de 12 ítems el error estándar de diseño es 3,96%, valor inferior al máximo permitido y un parámetro para establecer la calidad de la prueba.

En conclusión, después de realizar la validez discriminante y paralelamente el cálculo de la confiabilidad, esta investigación propone un cuestionario constituido por 12 ítems que es resultado de los primeros procesos de perfeccionamiento psicométrico que se le realizan a este instrumento de medición.

CONCLUSIONES

El cuestionario demostró validez facial por su claridad y facilidad de comprensión de cada uno de los ítems tanto para las expertas como para las mujeres que participaron en el estudio.

El cuestionario presentó un índice de validez de contenido general de 0,88, mayor que el mínimo valor exigido de 0,62, según Lawshe (13) para la prueba con 10 expertos, lo cual garantiza la validez de contenido desde la literatura científica relacionada con las prácticas de cuidado que realizan consigo mismas las mujeres en el posparto.

Al realizar la validez de escala se evidenció que es una escala corta y que queda para comprobación, posterior a futuros análisis según expertos y población de estudio, contrastándola con una teoría disciplinar.

La validez discriminante determinó los ítems que tenían mayor correlación y agrupó los que tenían correlaciones bajas; por tanto, al constituirse en una aproximación a la validez de constructo, permitió evidenciar la necesidad de seguir en el análisis de los ítems que se retiraron del cuestionario como propuesta de mejoramiento.

La direccionalidad de los reactivos es un mecanismo de aproximación al análisis de calidad de una prueba, lo que se evidenció en la dimensión de alimentación, en la cual todos sus ítems estaban enunciados en una misma dirección y, al final de los análisis estadísticos, la mayoría de los reactivos se conservaron porque presentaron alto poder de discriminación y correlación interna. La no

consideración de este aspecto lleva a graves consecuencias en la interpretabilidad de los resultados.

La confiabilidad determinada por el alfa de Cronbach en el cuestionario de 32 ítems presentó un alfa de Cronbach de 0,59 y en el de 12 ítems un alfa de Cronbach de 0,68, considerándose un puntaje aceptable y bueno en los primeros estudios de validación de un instrumento, según Nunnally y Bernstein (28). Aquí se confirma que el cuestionario tiene reactivos que miden lo mismo y que, contrario a lo que se considera en la teoría clásica, acortar la prueba hace que la confiabilidad aumente.

La confiabilidad del cuestionario permea todos los procesos de validación, por lo que su interpretación debe ir de la mano del perfeccionamiento continuo de la validez, puesto que un instrumento puede ser confiable y no necesariamente válido, pero para que un instrumento sea realmente confiable las respuestas repetidas deben surgir de lo que se pretende medir.

El cuestionario "Prácticas de cuidado que realizan consigo mismas las mujeres en el posparto" se constituye en el primer instrumento de medición aplicado a mujeres lactantes de diferentes estratos socioeconómicos y que se encontraban desde el primer día hasta los 180 días después del parto, para ser sometido a diferentes tipos de validez y comprobar si mide lo que se pretende medir, lo cual evidenció la importancia de seguir en el proceso de perfeccionamiento. Además, le permitió a la investigadora brindar educación sobre prácticas de cuidado en el posparto a las 176 mujeres que participaron en el estudio.

CUESTIONARIO
PRÁCTICAS DE CUIDADO QUE REALIZAN CONSIGO MISMAS
LAS MUJERES EN EL POSPARTO

Fecha: _____ No. consecutivo: _____
 Estrato: _____ Días posparto: _____
 Edad: _____ G _____ P _____ A _____ C _____

A continuación usted encuentra una serie de preguntas relacionadas con las prácticas de cuidado que generalmente realizan las mujeres en el posparto.

Marque con una X en la columna correspondiente la respuesta que refleje su situación:

Prácticas de cuidado	Nunca	Algunas veces	Siempre
1. ¿Consume al día mínimo 2 porciones, bien sea de carne, pollo, huevo, pescado o vísceras?			
2. ¿Consume al día mínimo 2 porciones de hortalizas o verduras (como zanahoria, berenjena, tomate, coliflor, remolacha, apio, lechuga, repollo, acelga, espinaca, pepino)?			
3. ¿Consume al día mínimo 3 o 4 veces leche o sus derivados (como queso, kumis, yogurt, cuajada)?			
4. ¿En su alimentación diaria las frutas están presentes?			
5. ¿Toma al día entre 7 y 8 vasos de líquido?			
6. ¿Se cepilla los dientes mínimo tres veces al día?			
7. ¿Acostumbra dialogar con su pareja o con alguna persona cercana a usted?			
8. ¿Busca apoyo en situaciones difíciles?			
9. ¿Consume hierro formulado por personal de salud?			
10. ¿Asistió o asistirá al control posparto?			
11. ¿Cuando se enferma acude a los servicios de salud?			
12. ¿Cumple con las recomendaciones sobre el cuidado dadas por el personal de salud?			

REFERENCIAS BIBLIOGRÁFICAS

- (1) Hernández R, Fernández C, Baptista P. Metodología de la investigación. 2^a. ed. México: McGraw-Hill; 2001, pp. 235-236.
- (2) Aguirre DC. Evaluación de la validez y confiabilidad de escalas usadas en salud. XII Simposio Nacional de Estadística de la Universidad Nacional de Colombia, Estadística en Ciencias de la salud, Armenia, 2003, p. 1.
- (3) Muñoz de Rodríguez L. Línea de investigación Prácticas de Cuidado Materno Perinatal. El arte y la ciencia del cuidado. Grupo de Cuidado, Facultad de Enfermería. Bogotá: Universidad Nacional de Colombia; 2002, p. 364.
- (4) Duhachek A, Coughlan A, Iacobucci D. Results on the standard error of the coefficient alpha index of reliability. Marketing Science. 2005; 24(2):294-301.
- (5) Khoja S, Scott RE, Casebeer AL, Mohsin M, Ishaq AF, Gilani S. Testing reliability of eHealth readiness assessment tools for developing countries. eHealth International Journal. 2007: 31-37.
- (6) Bonett D. Simple size requirements for testing and estimating coefficient alpha. Journal of educational and behavioral statistics, 2002. 27(4):335-340.
- (7) Nunnally J. Teoría psicométrica. 1^a. ed. México: Trillas; 1987, p. 310.

- (8) Balanza S, Morales I, Guerrero J, Conesa A. Fiabilidad y validez de un cuestionario para medir en estudiantes universitarios la asociación de la ansiedad y depresión con factores académicos y psicofamiliares durante el curso 2004-2005. *Revista Española de Salud Pública*. 2008; 28(2):192.
- (9) Norman G, Streiner DL. *Bioestadística*. Barcelona: Mosby; 1996, pp. 129-148.
- (10) Tavella N. Análisis de los ítems en la construcción de instrumentos psicométricos. México: Trillas; 1978, p. 75.
- (11) Herran O, Ardila M. Validez factorial del cuestionario de expectativas hacia el consumo de alcohol (AEQ-III) en adultos de Bucaramanga, Colombia. *Revista Médica de Chile*. 2007; 135(4):488-495.
- (12) Polit DF, Hungler BP. *Investigación científica en Ciencias de la Salud: principios y métodos*. 6ª. ed. México: McGraw-Hill Interamericana; 2000, pp. 398-401.
- (13) Lawshe CH. Quantitative approach to content validity. In: *Personnel Psychology*. 1975; 28:568.
- (14) Tristán López A. Elementos de diseño y análisis de instrumentos objetivos. Instituto de Evaluación e Ingeniería Avanzada, S. C. Material para uso interno del posgrado de Enfermería, Universidad Nacional de Colombia; 2008, pp. 2, 6, 8, 9, 25, 26, 27, 30.
- (15) Tristán A, Vidal R. Linear model to assess the scale's validity of a test. México; 2007, p. 1.
- (16) Grady K, Wallston B. *Research in Health Care Settings*. 1990; 2(1):88-89.
- (17) Hernández R, Fernández C, Baptista P. *Metodología de la investigación*. 4ª. ed. México: McGraw-Hill; 2006, pp. 439-442.
- (18) Ruiz Bolívar C. *Confiabilidad*. Programa Interinstitucional Doctorado en Educación. Universidad Pedagógica Experimental Libertador, Venezuela [consultado 18 de mayo de 2009]. En: <http://www.carlosruizbolivar.com/articulos/archivos/Curso%20CII%20%20UCLA%20Art.%20Confiabilidad.pdf>
- (19) Burns N, Grove S. *Investigación en enfermería*. 3ª. ed. España: Elsevier; 2004, p. 340.
- (20) Sturmey P, Newton JT, Cowley A, Bouras N, Holt G. The PAS-ADD checklist: Independent replication of its psychometric properties in a community sample. *British Journal of Psychiatry*, p. 319. 2005 [consultado 18 de mayo de 2009]. En: <http://bjp.rcpsych.org/cgi/content/full/186/4/319>
- (21) Barrera R, Reyes MC. Análisis comparado de las escalas de medición de la calidad de servicio. Actas de las XIII Jornadas Hispano-Lusas. La empresa familiar en un mundo globalizado. Lugo. Universidad de Santiago de Compostela. Facultad de Ciencias Económicas y Empresariales, Sevilla 2003 [consultado 18 de mayo de 2009]. En: http://www4.usc.es/Lugo-XIII-Hispano-Lusas/pdf/06_COMERCIALIZACION/28_barrera_reyes.pdf
- (22) Godoy-Izquierdo D, Godoy JF, López I, Martínez A, Gutiérrez S, Vázquez L. Propiedades psicométricas de la escala de autoeficacia para el afrontamiento del estrés (EAEAE). *Psicothema*. 2008; 20(1):161 [consultado 18 de mayo de 2009] En: <http://dialnet.unirioja.es/servlet/articulo?codigo=2519148>
- (23) Klassen PT. New TTI instrument reliability studies. Target Training International, Ltd., pp. 1-2, 2003 [consultado 18 de mayo de 2009]. En: <http://www.nielsongroup.com/hiring-forfit/TriMetricTTIValidity.pdf>
- (24) Soler S, Rodríguez R, Fernández B, Moreno M. Análisis de confiabilidad del test de habilidades múltiples para el ingreso a las carreras de ciencias médicas. *Escuela Nacional de Salud Pública. Revista Cubana Educación Médica Superior*; 16(4):2 [consultado 18 de mayo de 2009]; En: http://bvs.sld.cu/revistas/ems/vol16_4_02/ems03402.htm
- (25) Canaval GE. Propiedades psicométricas de una escala para medir percepción del empoderamiento comunitario en mujeres. *Revista Colombia Médica*. 1999; 30(2): 72 [consultado 10 de febrero de 2008]. En: <http://redalyc.uaemex.mx/redalyc/pdf/283/28330203.pdf>
- (26) Llerena M. Metodología para la evaluación de la calidad de estrategias didácticas de cursos a distancia (MACCAD). *Formación Universitaria*. 2008; 1(2):45 [consultado 18 de mayo de 2009]. En: <http://www.citchile.cl/revista-formacion/v1n2fu/art06.pdf>
- (27) Ruiz CH, Fajardo MT, Velandia JP. Validez y confiabilidad del instrumento Prácticas de cuidado que realizan consigo mismas y con el hijo por nacer las mujeres en la etapa prenatal. *Avances en Enfermería*. 2006; 24(1):31-32.
- (28) Nunnally JC, Bernstein I. *Teoría psicométrica*. 3ª. ed. México: McGraw-Hill; 1995, pp. 209-281.
- (29) Cipagauta S, Corzo JC. Validación de una escala para medir actitudes hacia los factores de la cultura organizacional en ambientes de cambio. Tesis de grado, Universidad Nacional de Colombia, Facultad de Ciencias Humanas, Departamento de Psicología; 2004, pp. 142-155.
- (30) Ledesma R, Molina G, Valero P. Análisis de consistencia interna mediante alfa de Cronbach: un programa basado en gráficos dinámicos. *Revista Psico-USF*. 2002; 7(2): 143-152.
- (31) Gliem JA, Gliem RR. Calculating, interpreting, and reporting Cronbach's Alpha reliability coefficient for Likert-Type scales. Midwest research to practice conference in adult, continuing, and community education, pp. 82-88 [consultado 20 de mayo de 2009]. En: <https://idea.iupui.edu/dspace/handle/1805/344>.