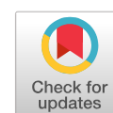


Revista Española de Nutrición Humana y Dietética

Spanish Journal of Human Nutrition and Dietetics

www.renhyd.org

ARTÍCULO DE REVISIÓN – INFORMACIÓN SUPLEMENTARIA

Recomendaciones para la validación de instrumentos que evalúan conducta alimentaria: una revisión narrativa

> Recommendations for the validation of instruments assessing eating behavior: a narrative review

Claudia Hunot Alexander^a, Leyna Priscila López-Torres^{a,b*}, Patricia Curiel-Curiel^a, Jocelyn González Toribio^a

^a Instituto de Nutrición Humana, Centro Universitario de Ciencias de la Salud de la Universidad de Guadalajara. Guadalajara 44340, Jalisco, México

^b Departamento de Ciencias Sociales, Centro Universitario de Ciencias de la Salud de la Universidad de Guadalajara. Guadalajara 44340, Jalisco, México;

* priscila.lopez@academicos.udg.mx

Recibido: 02/03/2026; Aceptado: 08/05/2026; Publicado: 12/08/2026

Editora asignada: Rafael Almendra-Pegueros, Departamento de Ciencias Básicas de la Salud, Facultad de Ciencias de la Salud, Universidad Rey Juan Carlos (URJC), Alcorcón, Madrid, España.

CITA

Hunot Alexander C, López-Torres LP, Curiel-Curiela P, González Toribio J, Vaqué-Crusellas C, Ribot B, Rierola-Fochs S, Vila-Martí A, Torres-Moreno M, Merchán Baeza JA. Recomendaciones para la validación de instrumentos que evalúan conducta alimentaria: una revisión narrativa. Rev Esp Nutr Hum Diet. 2026; 30(4), e2783 doi: 10.14306/renhyd.30.4.2783



Anexo Suplementario S1. Definiciones técnicas operativas, puntos de corte e implicaciones metodológicas

1. Adaptación transcultural (traducción–retrotraducción)

Proceso sistemático mediante el cual un instrumento desarrollado en un idioma o contexto cultural se traduce y adapta para preservar equivalencia semántica, conceptual y cultural en una nueva población^{1–4}. La adaptación lingüística no equivale a validación completa. Debe complementarse con evaluación psicométrica en la nueva muestra.

Procedimiento recomendado

- Traducción directa independiente (≥ 2 traductores)
- Síntesis
- Retrotraducción
- Comité de expertos
- Pilotaje cognitivo
- Versión final

Puntos críticos

No existen “puntos de corte”, pero sí estándares mínimos:

- Documentar equivalencia semántica y conceptual.
- Reportar discrepancias y decisiones del comité.

2. Validez de contenido

Grado en que los ítems representan adecuadamente el dominio de la dimensión o constructo^{2,3}. Debe definirse número de expertos, criterios de selección y reglas de decisión.

Procedimientos

- Panel de expertos
- Método Delphi⁵
- Índice de Validez de Contenido por Ítem (*Item-level Content Validity Index*, I-CVI)
- Índice de Validez de Contenido a Nivel de Escala (*Scale-level Content Validity Index*, S-CVI)^{5,6}

Puntos de corte comúnmente reportados

- I-CVI ≥ 0.78 (≥ 6 expertos)
- S-CVI ≥ 0.80 – 0.90

3. Validez aparente (*Face validity*)

Percepción subjetiva de que el instrumento mide lo que pretende medir^{1,2,4}. No constituye evidencia suficiente por sí sola.

Evaluación

- Juicio de expertos
- Retroalimentación de participantes

4. Evidencia basada en el proceso de respuesta

Análisis de los procesos cognitivos que utilizan los participantes al responder los ítems^{1,2,7,8}. Procesos clave para asegurar equivalencia conceptual en adaptaciones transculturales.

Métodos

- Entrevistas cognitivas⁷
- Técnica “*think aloud*”^{8,9}
- Análisis de interpretación semántica

5. Validez de constructo: estructura interna

Se refiere al grado en que la estructura factorial de un instrumento, evaluada mediante análisis como el Análisis de Componentes Principales (ACP), Análisis Factorial Exploratorio (AFE) y el Análisis Factorial Confirmatorio (AFC), es consistente con la teoría que sustenta el constructo que se pretende medir^{1,2,10–12}.

5.1 Análisis de Componentes Principales (*Principal Component Analysis, PCA/ACP*)

Técnica de reducción de dimensionalidad busca identificar factores latentes que explican la covariación entre los ítems, por lo que resulta más adecuado cuando el objetivo es explorar o depurar la estructura de un constructo psicológico o conductual^{1,2,12}. Útil para derivar patrones de alimentación^{13,14}.

Cargas factoriales

No existen puntos de corte universales para cargas factoriales, pero se recomienda¹²:

- Cargas ≥ 0.30 mínimas
- ≥ 0.40 – 0.50 deseables

Requisitos

- Kaiser–Meyer–Olkin (KMO) ≥ 0.60
- Prueba de esfericidad de Bartlett significativa; entre más pequeño el valor, mejor.
- La selección del número de factores puede apoyarse en el criterio de Kaiser (eigenvalues ≥ 1.0), la inspección del gráfico de sedimentación (scree plot) y, preferentemente, el análisis paralelo, considerado más robusto.
- La elección del tipo de rotación debe justificarse teóricamente: las rotaciones ortogonales (p. ej., Varimax) asumen independencia entre factores, mientras que las oblicuas (p. ej., Oblimin o Promax) permiten correlaciones interfactoriales, siendo estas últimas generalmente más apropiadas en constructos psicológicos.

5.2 Análisis Factorial Exploratorio (*Exploratory Factor Analysis, EFA/AFE*)

Explora la estructura latente que explican la covariación entre los ítems, cuando no existe modelo confirmado^{11,12}.

Cargas factoriales

- ≥ 0.40 aceptable
- ≥ 0.50 fuerte

Requisitos

- Iguales al PCA/ACP

5.3 Análisis Factorial Confirmatorio (*Confirmatory Factor Analysis, CFA/AFC*)

Contrasta ajuste de modelo teórico predefinido. Sirve para replicar la estructura factorial en una segunda muestra, mostrando una estrategia secuencial útil cuando se combina exploración y confirmación^{11,15}.

Requisitos

- El AFC permite evaluar qué tan bien se ajusta el modelo a los datos mediante índices de ajuste como el Índice de ajuste comparativo^{1,2,12,15}.

Índices de ajuste comunes

Índice	Punto de corte comúnmente reportado
Comparative Fit Index (CFI)	≥ 0.90 aceptable; ≥ 0.95 excelente
Tucker-Lewis Index (TLI)	≥ 0.90 aceptable; ≥ 0.95 excelente
Root Mean Square Error of Approximation (RMSEA)	≤ 0.08 aceptable; ≤ 0.06 óptimo
Standardized Root Mean Square Residual (SRMR)	≤ 0.08

Fuente: ^{1,2,12,15}

- En el AFC también es relevante examinar los **índices de modificación** y los **residuales estandarizados**, que permiten identificar posibles áreas de mal ajuste del modelo. Los índices de modificación estiman la mejora esperada en el ajuste si se libera un parámetro restringido (por ejemplo, correlación entre errores), mientras que los residuales indican discrepancias entre la matriz observada y la estimada. No obstante, cualquier reespecificación del modelo debe justificarse teóricamente y no basarse únicamente en criterios estadísticos^{10,11,15,16}

- Se utilizan valores más bajos del Criterio de Información de Akaike (*Akaike's Information Criterion*, AIC) y del Criterio de Información Bayesiano (*Bayesian Information Criterion*, BIC) como indicadores de mejor ajuste relativo entre modelos comparados. Ambos criterios penalizan la complejidad del modelo (número de parámetros), por lo que un valor menor sugiere un modelo que logra un mejor equilibrio entre ajuste y parsimonia^{11,17}

5.4 Invariancia de medición (*Measurement Invariance*)

La invariancia de medición evalúa si un instrumento mide el mismo constructo de la misma manera en diferentes grupos (por ejemplo, sexo, edad, cultura) o a través del tiempo. Es un requisito previo para realizar comparaciones válidas entre grupos o para interpretar cambios longitudinales^{2,12}.

La evaluación de la invariancia se realiza generalmente mediante modelos jerárquicos dentro del marco del Análisis Factorial Confirmatorio (*Confirmatory Factor Analysis*, CFA/AFC) o del Modelo de Ecuaciones Estructurales (*Structural Equation Modelling*, SEM).

Tipos de invariancia

1. Invariancia configural

- Evalúa si la estructura factorial (número de factores y patrón de cargas) es equivalente entre grupos.
- No impone restricciones en los parámetros.
- Es el nivel mínimo necesario.

2. Invariancia métrica (o factorial débil)

- Restringe las cargas factoriales a ser iguales entre grupos.
- Permite comparar relaciones entre variables latentes.

3. Invariancia escalar (o fuerte)

- Restringe cargas factoriales e interceptos.
- Permite comparar medias latentes entre grupos.

4. Invariancia estricta (o residual)

- Añade restricción de varianzas de error.
- Permite comparación más estricta de puntuaciones observadas.

Criterios de evaluación

Tradicionalmente se utilizaba la prueba de diferencia de chi-cuadrado ($\Delta\chi^2$), pero dado que es sensible al tamaño muestral, actualmente se recomienda evaluar cambios en índices de ajuste:

- $\Delta CFI \leq 0.01$
- $\Delta RMSEA \leq 0.015$
- $\Delta SRMR \leq 0.030$ (métrica) o ≤ 0.015 (escalar)

Implicaciones metodológicas

- Si no se demuestra al menos invariancia métrica, no deben compararse asociaciones entre grupos.
- Si no se alcanza invariancia escalar, no deben compararse medias latentes.
- En ausencia de invariancia completa, puede explorarse invariancia parcial¹⁰

5.5 Modelos de Ecuaciones Estructurales (*Structural Equation Modelling, SEM*)

Extensión del CFA que permite modelar relaciones entre variables latentes. Pueden ampliar el análisis de estructura interna hacia la evaluación de relaciones entre constructos latentes, siempre que se cuente con hipótesis teóricas claras y tamaño muestral suficiente^{2,10}.

Requiere:

- Tamaño muestral adecuado (≥ 200 recomendado)
- Justificación teórica

6. Validez de constructo: relación con otras variables

Validez de constructo basada en la relación con otras variables se refiere al grado en que las puntuaciones de un instrumento se asocian con otras medidas o variables de manera consistente con las hipótesis teóricas del constructo evaluado^{1,2}.

Incluye:

- Evidencia convergente (correlaciones con constructos afines)
- Evidencia discriminante (bajas correlaciones con constructos distintos)
- Comparación entre grupos conocidos

Correlaciones esperadas¹⁸:

- $r \geq 0.30$ moderadas
- $r \geq 0.50$ fuertes

7. Validez de criterio

Grado en que las puntuaciones de un instrumento se relacionan con un criterio externo considerado como referencia o estándar, ya sea medido simultáneamente o en un momento posterior^{1,2}.

7.1 Validez de criterio: Concurrente

Comparación simultánea con criterio externo (ej. entrevista clínica)^{1,2}.

Coeficientes:

- $r \geq 0.50$ fuerte

7.2 Validez de criterio: Predictiva

Capacidad para anticipar desenlaces futuros^{1,2}.

Requisitos:

- Diseño longitudinal
- Seguimiento temporal

8. Confiabilidad: consistencia interna

La consistencia interna se refiere al grado en que los ítems de un instrumento están interrelacionados y miden de manera homogénea el mismo constructo. Se evalúa comúnmente mediante el coeficiente alfa de Cronbach (α) y el coeficiente omega de McDonald (ω), siendo este último preferible cuando no se cumple el supuesto de tau-equivalencia^{1,2}.

8.1 Alfa de Cronbach (*Cronbach's alpha*)

Es el índice más reportado, aunque su interpretación supone condiciones como la tau-equivalencia esencial (cargas factoriales similares entre ítems) ^{1,2,12}. De manera clásica, se consideran aceptables valores ≥ 0.70 en contextos exploratorios, con exigencias mayores para usos clínicos o decisiones individuales^{1,2}.

Requisitos:

- Supone tau-equivalencia.
- **Puntos de corte** comunes, con exigencias mayores para usos clínicos o decisiones individuales^{1,12}:
 - ≥ 0.70 aceptable
 - ≥ 0.80 buena
 - ≥ 0.90 excelente

8.2 Omega de McDonald (*McDonald's omega*)

Adecuado en modelos congénicos (cargas desiguales) y estructuras más complejas, especialmente cuando existe evidencia de estructura interna mediante AFE/AFC¹⁹.

Puntos de corte

Interpretación similar a alfa:

- ≥ 0.70 recomendable

9. Confiabilidad: estabilidad temporal

La estabilidad temporal evalúa la reproducibilidad de los puntajes cuando no se espera un cambio real del constructo. Suele estimarse mediante test-retest y, cuando se busca concordancia entre mediciones repetidas, con el coeficiente de correlación intraclase (CCI)^{12,20}.

Test-retest

Coeficiente de correlación intraclase (*Intraclass Correlation Coefficient, ICC/CCI*)

Puntos de corte comunes:

- < 0.50 pobre
- 0.50–0.75 moderada
- 0.75–0.90 buena
- 0.90 excelente

Referencias

- (1) Furr RM Psychometrics: an introduction. 4th. ed. Los Angeles London New Delhi Singapore Washington, DC Melbourne: SAGE Publications; 2022.
- (2) Streiner DL, Norman GR, Cairney John Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford University Press; 2024.
- (3) Boateng GO, Neilands TB, Frongillo EA, Melgar-Quinonez HR, Young SL Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Front Public Health*. 2018;6:149, doi: 10.3389/fpubh.2018.00149.
- (4) World Health Organization (WHO) Global Scales for Early Development v1.0. Adaptation and translation guide. World Health Organization; 2023.
- (5) Hsu C-C, Sandford BA The Delphi Technique: Making Sense of Consensus. *Pract Assess Res Eval*. 2007;12(1), doi: 10.7275/pdz9-th90.
- (6) Lynn MR Determination and Quantification Of Content Validity. *Nurs Res*. 1986;35(6):382.
- (7) Banna JC, Vera Becerra LE, Kaiser LL, Townsend MS Using Qualitative Methods to Improve Questionnaires for Spanish Speakers: Assessing Face Validity of a Food Behavior Checklist. *J Am Diet Assoc*. 2010;110(1):80-90, doi: 10.1016/j.jada.2009.10.002.
- (8) Fox MC, Ericsson KA, Best R Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychol Bull*. 2011;137(2):316-44, doi: 10.1037/a0021663.
- (9) Eccles DW, Aarsal G The think aloud method: what is it and how do I use it? *Qual Res Sport Exerc Health*. 2017;9(4):514-31, doi: 10.1080/2159676X.2017.1331501.
- (10) Kline RB Principles and Practice of Structural Equation Modeling. Guilford Publications; 2023.
- (11) Dugard P, Todman J, Staines H Factor analysis. Approaching multivariate analysis. London and New York: Routledge; 2010. p. 177-205.
- (12) Field A Discovering Statistics Using IBM SPSS Statistics. SAGE Publications; 2024.
- (13) Gil-Barrera AS, Hunot-Alexander C, Chávez-Palencia C, González-Toribio J, Casillas-Toral E, Álvarez-Zaragoza DC, et al. Appetitive Traits and Dietary Patterns in Mexican Children Aged 12 to 36 Months. *Nutrients*. 2025;17(11), doi: 10.3390/nu17111814.
- (14) Denova-Gutiérrez E, Ramírez-Silva I, Rodríguez-Ramírez S, Jiménez-Aguilar A, Shamah-Levy T, Rivera-Dommarco JA, et al. Validity of a food frequency questionnaire to assess food intake in Mexican adolescent and adult population. *Salud Pública México*. 2016;58(6):617-28, doi: 10.21149/spm.v58i6.7862.
- (15) Hu, Li-tze., Bentler, Peter M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model Multidiscip J*. s. f.;6(1), doi: <https://doi.org/10.1080/10705519909540118>.

- (16) Marsh HW, Muthén B, Asparouhov T, Lüdtke O, Robitzsch A, Morin AJS, et al. Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching. *Struct Equ Model Multidiscip J*. 2009;16(3):439-76, doi: 10.1080/10705510903008220.
- (17) Hunot C, Fildes A, Croker H, Llewellyn CH, Wardle J, Beeken RJ Appetitive traits and relationships with BMI in adults: Development of the Adult Eating Behaviour Questionnaire. *Appetite*. 2016;105:356-63, doi: 10.1016/j.appet.2016.05.024.
- (18) Cohen J Statistical Power Analysis. *Curr Dir Psychol Sci*. 1992;1(3):98-101, doi: 10.1111/1467-8721.ep10768783.
- (19) DeVellis RF, Thorpe CT Scale Development: Theory and Applications. SAGE Publications; 2021.
- (20) Viladrich C, Angulo-Brunet A, Doval E Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *An Psicol*. 2017;33(3):755, doi: 10.6018/analesps.33.3.268401.

Anexo Suplementario S2

Anexo S2.1. Guía operativa para desarrollo, adaptación y validación de instrumentos de conducta alimentaria

A.1 Desarrollo de un instrumento nuevo

Definición

- Definir la(s) **dimensión(es)** (patrones de alimentación / procesos psicológicos, emocionales y sociales / conductas alimentarias disfuncionales)
- Especificar propósito (investigación, clínica, tamizaje, seguimiento) y población

1. Generación y depuración inicial de ítems

- Emplear un marco teórico y mapa de dominios
- Generar los ítems necesarios (literatura, entrevistas, expertos)
- Ajustar la redacción (lenguaje, periodo de referencia, opciones de respuesta)

2. Evidencia temprana

- Elegir los métodos de validez de contenido (panel de expertos; si aplica Delphi; I-CVI/S-CVI)
- Explorar la validez aparente (claridad/aceptabilidad en población objetivo)
- Valorar el proceso de respuesta (entrevistas cognitivas / *"think aloud"*; ajustes derivados)

3. Validez de constructo: Estructura interna

- Realizar el Análisis Factorial Exploratorio (AFE) o Análisis Factorial Confirmatorio (AFC), dependiendo del objetivo del estudio) y realizar depuración de ítems según carga factorial.
- Realizar la confirmación del modelo a través del Análisis Factorial Confirmatorio (AFC) e índices de ajuste y, cuando es pertinente, índices de modificación y residuales.
- Realizar Invariancia cuando el objetivo incluye comparaciones entre grupos (por ejemplo, sexo, edad o contextos culturales) (métrica: igualdad de cargas factoriales; escalar: igualdad de interceptos/umbrales; y, residual: igualdad de errores de medición).

4. Relación con otras variables

- Contrastar la hipótesis a priori: convergente/discriminante/grupos conocidos
- Calcular tamaños de efecto

5. Confiabilidad

- Determinar la consistencia interna (alfa Cronbach y/o omega McDonald por subescala)
- Medir la estabilidad temporal (test-retest / ICC; intervalo justificado)

6. Criterio

- Determinar la validez concurrente (si existe criterio externo pertinente)
- Determinar la validez predictiva (si hay seguimiento y desenlace definido)

7. Interpretación y reporte

- Determinar la puntuación, subescalas, manejo de faltantes, puntos de corte
- Identificar las limitaciones y uso recomendado (investigación vs clínica)

Preparación

- Confirmar autorización y versión fuente
- Definir población/variante del español y equivalencia conceptual

1. Traducción y revisión

- Realizar traducción directa (≥ 2) + reconciliación
- Realizar retrotraducción independiente
- Colaborar con comité (temático + lingüístico) y hacer registro de decisiones

2. Pilotaje y comprensión

- Realizar la validez aparente (aceptabilidad/claridad)
- Realizar el proceso de respuesta (entrevistas cognitivas/"*think aloud*"; cambios)

3. Evaluación psicométrica en la población destino

- Determinar la estructura interna (ACP/AFE y ACP para explorar si hay cambios culturales)
- Determinar la invariancia (si se comparará con grupos/otras versiones)
- Determinar la relación con otras variables (hipótesis a priori)
- Determinar la confiabilidad (alfa de Cronbach/omega de McDoland; test-retest/ICC)

4. Criterio (si aplica)

- Realizar la validez concurrente con criterio externo (más típico en tamizaje clínico)
- Realizar la validez predictiva (si hay seguimiento)

A.3 Validación de un instrumento existente en una nueva muestra (misma lengua)

- Justificar transferibilidad (población/edad/contexto)
- Confirmar estructura interna (ACP/AFE y ACP) y, si se realiza comparación entre grupos, invarianza
- Determinar la relación con otras variables (hipótesis a priori)
- Calcular la confiabilidad (alfa Cronbach/omega McDonald; test-retest/ICC si procede)
- Evaluar validez de criterio (concurrente o predictiva) cuando exista un estándar externo pertinente al constructo)
- Reportar pormenorizadamente las limitaciones: "validación en muestra X" $\neq$ "validez universal"

Anexo S2.2. Diagrama de flujo. Proceso para el desarrollo, adaptación o validación de instrumentos de conducta alimentaria

Paso 1. Definir el escenario metodológico ¿Qué se desea hacer?		
A. Desarrollar un instrumento nuevo	B. Adaptar transculturalmente un instrumento existente	C. Validar un instrumento existente en una nueva muestra (misma lengua)
Siempre verificar: • Coherencia conceptual • Tamaño muestral adecuado • Justificación teórica de decisiones estadísticas • Correspondencia entre propósito y evidencia generada		
1. Delimitación conceptual • Definir dimensión(es): ○ Patrones de alimentación ○ Procesos psicológicos, emocionales y sociales ○ Conductas alimentarias disfuncionales • Especificar propósito (investigación, clínica, tamizaje, seguimiento) • Definir población objetivo	1. Confirmar versión fuente y autorización • Verificar versión original actualizada • Contar con permiso formal de uso	1. Justificar transferibilidad • Fundamentar la aplicabilidad del instrumento a la nueva población (edad/contexto)
2. Generación de ítems • Marco teórico y mapa de dominios • Revisión de literatura / entrevistas / expertos • Redacción clara y periodo de referencia definido	2. Traducción–retrotraducción • Traducción directa (≥ 2) • Reconciliación • Retrotraducción • Comité experto	2. Confirmar estructura interna • AFC preferente • Invariancia si hay comparación entre grupos •

3. Evidencia preliminar

- Validez de contenido (panel / I-CVI / S-CVI)
- Validez aparente
- Proceso de respuesta (entrevistas cognitivas)

3. Pilotaje

- Validez aparente
- Entrevistas cognitivas

3. Relación con otras variables

- Hipótesis explícitas

4. Estructura interna

- AFE (explorar) o AFC (explorar)
- ACP (confirmar modelo teórico)
- Índices de ajuste
- Índices de modificación (si aplica)
- Invariancia (si habrá comparación entre grupos)

4. Evaluación psicométrica

- Confirmar estructura interna
 - AFE (explorar) o AFC (explorar)
 - ACP (confirmar modelo teórico)
- Invariancia (si se comparará con versión original)
- Relación con otras variables
- Confiabilidad

4. Confiabilidad

- α / ω
- ICC si procede

5. Relación con otras variables

- Hipótesis a priori
- Evidencia convergente/discriminante
- Tamaños de efecto

5. Criterio (si aplica)

- Concurrente
- Predictiva

5. Criterio (si existe estándar externo pertinente)**6. Confiabilidad**

- Consistencia interna (α / ω)
- Estabilidad temporal (ICC / test-retest)

6. Reportar

- Equivalencia conceptual
- Limitaciones

7. Criterio (si aplica)

- Concurrente
- Predictiva

8. Interpretación y reporte

- Sistema de puntuación
 - Subescalas
 - Manejo de faltantes
 - Limitaciones
-