

Nivel de cumplimiento de procesos de validación y análisis de escalas de medida en pediatría

Level of compliance with validation processes and analysis of measurement scales in pediatrics

José A. Martínez^a, Laura Martínez^a

^aFacultad de Ciencias de la Empresa, Universidad Politécnica de Cartagena. Cartagena, España.

Recibido: 28 de diciembre de 2023; Aceptado: 6 de mayo de 2024

¿Qué se sabe del tema que trata este estudio?

El uso de escalas de medida de constructos latentes es ubicuo en la literatura médica. La necesidad de que el instrumento de medida sea válido es condición esencial para su correcta aplicación, siendo deseable, asimismo, que tenga una alta fiabilidad.

¿Qué aporta este estudio a lo ya conocido?

El nivel de cumplimiento de premisas metodológicas y estadísticas de los estudios de validación de instrumentos psicométricos publicados en dos revistas de pediatría en idioma español. Entrega recomendaciones para futuras investigaciones.

Resumen

El uso de escalas de medida de constructos latentes es ubicuo en la literatura médica. Una condición esencial es que dicho instrumento sea validado, siendo deseable, además, que tenga una alta fiabilidad. **Objetivo:** Evaluar el nivel de cumplimiento de diversas premisas metodológicas y estadísticas de estudios de validación de instrumentos psicométricos en el ámbito de la pediatría, necesarias para garantizar una adecuada interpretación de los resultados, y ofrecer recomendaciones para futuras investigaciones. **Método:** Se identificaron los artículos publicados en dos de las revistas más prestigiosas en idioma español, Andes Pediátrica y Anales de Pediatría entre diciembre de 2000 diciembre de 2023, en cuyo método de análisis se realiza análisis factorial de los ítems de escalas de medida, y en los que se estudia la fiabilidad y validez de esas escalas. Se seleccionaron 25 artículos que atendían a las condiciones de la búsqueda, los cuales fueron contrastados con las recomendaciones de procedimiento de análisis. **Resultados:** Se identificaron 5 puntos clave sobre los que incidir en rigurosidad: 1) Indicadores reflectivos y formativos; 2) Fijación del mejor indicador reflectivo; 3) Validez de una escala ligada al ajuste del modelo teórico; 4) Fiabilidad usando el alfa de Cronbach; y 5) Análisis factorial frente a componentes principales. Rotación de factores. **Conclusiones:** Los procesos de análisis de instrumentos de medida en pediatría deben perfeccionarse, especialmente en lo relativo a la selección y especificación de indicadores, el uso del análisis factorial y el ajuste de los modelos planteados.

Palabras clave:

Escalas de Medidas;
Validez;
Fiabilidad;
Análisis Factorial;
Ajuste Exacto;
Ecuaciones
Estructurales

Abstract

The use of latent construct measurement scales is widespread in the medical literature. An essential condition is that the instrument be validated, in addition to being highly reliable. **Objective:** To evaluate the level of compliance with various methodological and statistical premises of validation studies of psychometric instruments in pediatrics, which are necessary to ensure adequate interpretation of the results, in order to offer recommendations for future research. **Methods:** Articles published in two of the most prestigious journals in Spanish (*Andes Pediátrica* and *Anales de Pediatría*) were identified between December 2000 and December 2023, where factor analysis was performed on the items of measurement scales, and in which the reliability and validity of these scales were studied. Twenty-five articles that met the selection criteria were selected and compared with the analysis procedure recommendations. **Results:** Five key points were identified to improve: 1) Reflective and formative indicators; 2) Determination of the best reflective indicator; 3) Validity of a scale linked to the fit of the theoretical model; 4) Reliability using Cronbach's alpha; 5) Factor analysis versus principal components. Factor rotation. **Conclusions:** The analysis processes of measurement instruments in pediatrics must be improved, especially regarding the selection and specification of indicators, the use of factor analysis, and the fit of the proposed models.

Keywords:

Measurement Scales;
Validity;
Reliability;
Factor Analysis;
Exact Fit;
Structural Equations

Introducción

El uso de escalas de medida de constructos latentes es ubicuo en ciencias sociales y humanas¹ incluyendo, por supuesto, la literatura médica. La necesidad de que el instrumento de medida sea válido es condición esencial para su correcta aplicación, siendo deseable, asimismo, que tenga una alta fiabilidad. Validez y fiabilidad son propiedades que se suelen testar con herramientas estadísticas, entre ellas el análisis factorial, tanto exploratorio como confirmatorio, siendo este último enmarcado en los modelos de ecuaciones estructurales.

Sin embargo, la interacción de estas propiedades y las herramientas para su estudio no es sencilla y, con frecuencia, está sujeta a incorrectas estrategias de análisis, propiciadas a veces por errores básicos de conceptualización, y en otras ocasiones por la dificultad de interpretar ciertas ambigüedades que la literatura sobre psicometría muestra.

Es obvio que una incorrecta metodología de análisis invalida los resultados y, por ende, las conclusiones de los estudios. Este hecho cobra especial relevancia en disciplinas como la medicina y, dentro de ella, la pediatría, donde las acciones basadas en conclusiones erróneas pueden producir un coste evidente.

El objetivo de este trabajo es evaluar el nivel de cumplimiento de varias premisas metodológicas y estadísticas en los estudios de validación de instrumentos psicométricos en el ámbito de la pediatría, que son necesarias para garantizar una adecuada interpretación de los resultados, ofreciendo así recomendaciones para futuras investigaciones.

Método

Se identificaron los artículos publicados en *Andes Pediátrica* y *Anales de Pediatría* entre diciembre de 2000 diciembre de 2023, en cuyo método de análisis se realiza análisis factorial de los ítems de escalas de medida, y en los que se estudia la fiabilidad y validez de esas escalas. Se encontraron 25 artículos²⁻²⁶ que atendían a las condiciones de la búsqueda, los cuales fueron contrastados con las recomendaciones de procedimiento de análisis que se describen a continuación.

Resultados

Indicadores reflectivos y formativos

El primer paso para la correcta especificación de un modelo es establecer la adecuada relación causal entre los indicadores observables y las variables latentes subyacentes. Así, si esa variación es causada por la variable latente se denominan indicadores reflectivos, y si es la variable latente la que se ve afectada por la variación en los indicadores se denominan indicadores formativos^{27,28}. Las ecuaciones [1] y [2] muestran, respectivamente, una especificación general de ambos tipos de indicadores.

$$y_j = \alpha_j + \lambda_j \eta + \theta_j \quad [1]$$

$$\eta = \sum_j \lambda_j^* y_j + \zeta \quad [2]$$

Donde y_j son los j indicadores observables, y η es la variable latente que pretenden medir, α_j es la orde-

nada en el origen, λ_j es la carga factorial, θ_j y ζ es ruido blanco. El asterisco significa que las cargas factoriales no tienen por qué coincidir con λ_j en [1]. Según el tipo de especificación formativa, ζ puede ser cero²⁷.

No es sencillo a veces la distinción, pero básicamente podemos afirmar que los indicadores reflectivos son intercambiables y deben correlacionar altamente, de forma que, si uno de ellos se elimina no altera el significado de la variable latente. Sin embargo, los indicadores formativos no son intercambiables, y no tienen por qué correlacionar. Si se elimina uno de los indicadores, se altera el significado de la variable latente, ya que esa variable se define a partir de esos indicadores. Los indicadores reflectivos son manifestaciones de la variable latente, y en el caso en el que todos se midan en la misma escala, con un solo indicador bastaría para estimar el valor de la variable latente. No ocurre así con los formativos, ya que el valor de todos ellos influye en el valor de la variable latente, por lo que esta cambia en función de si los indicadores lo hacen. La discusión y los detalles son más complejos²⁷, pero esta simplificación es un buen punto de partida.

Por ejemplo, en uno de los artículos publicados en Anales de Pediatría sobre la escala de dolor postoperatorio¹⁴ –Parent's Postoperative Pain Measurement (PPPM)–, habría que valorar si realmente los indicadores son formativos en lugar de reflectivos. Tras realizar un análisis de componentes principales se identifican 3 factores principales: expresiones de dolor, interferencia funcional y comportamiento del dolor. Cada uno de ellos está medido con varios ítems. Se podrían entonces plantear las siguientes preguntas; 1) ¿los 3 factores tienen que estar necesariamente correlacionados de forma importante? Parece plausible concebir que no tenga por qué ser así, ya que, por ejemplo, los indicadores relacionados con la conducta ante la comida podrían no estar fuertemente asociados a los relacionados con las expresiones de dolor; 2) ¿Se pueden intercambiar los ítems o eliminar alguno sin que varíe el significado de la variable? No lo parece. Por ejemplo, los ítems 11 y 15 del factor “comportamiento del dolor” son difícilmente intercambiables, ya que uno se refiere a si el niño golpea la parte dolorida de su cuerpo y el otro a si toma la medicación. Por tanto, se están midiendo conceptos diferentes; 3) ¿Una mayor puntuación en la escala global significa un valor más alto del constructo latente? Parece ser así, ya que cuanto más alto puntué en cada ítem mayor será el dolor postoperatorio. Es decir, un niño con valores altos en las expresiones de dolor, en la interferencia funcional y en el comportamiento del dolor, será considerado que tiene más dolor postoperatorio que un niño con valores altos en las expresiones de dolor, pero valores un poco más bajos en los otros dos factores.

De este modo, esta propuesta que hacen los autores

podría ser candidata a considerarse formativa en lugar de reflectiva. Las implicaciones que ello tiene para la modelización y el análisis estadístico son importantes²⁷.

Otro ejemplo análogo se encuentra en el estudio de una escala de evaluación funcional de extremidades superiores en niños con distrofia muscular¹⁸, planteándose 21 ítems con un puntaje total de la escala entre 0 y 120, siendo 0 el valor mínimo y 120 el valor máximo. En general, este tipo de escalas en las que se forma un índice sumativo cuyo valor gradúa el atributo que se quiere medir son formativas, no reflectivas. De este modo, no quedaría justificado realizar un análisis factorial exploratorio. Más casos similares se pueden ver en otras publicaciones¹⁹.

Estos mismos planteamientos habría que hacerlos para cada uno de los instrumentos de medición que aparecen publicados en Andes Pediatría y Anales de Pediatría de forma que quede bien justificada la naturaleza reflectiva o formativa del instrumento de medida, condicionando así el uso de herramientas como el alfa de Cronbach o el análisis factorial exploratorio y confirmatorio, como posteriormente se explicará.

Fijación del mejor indicador reflectivo

Si los indicadores son reflectivos, conviene fijar la varianza de error del mejor indicador posible²⁹. Esto hace que exista un compromiso por parte del investigador con el significado de la variable latente, que se manifiesta a través del mejor indicador observable. De este modo, el significado de la variable latente es robusto a los cambios derivados de lo que se conoce como *interpretational confounding*²⁷ (la traducción poco amigable podría ser “confusión interpretativa”), que básicamente se refiere al cambio en las cargas factoriales de los indicadores observables, es decir, en la relación entre la variable latente y los ítems que la miden, cuando cambia la estructura de covarianza de los datos, por ejemplo, al incluir otras variables en el modelo.

Evidentemente, esto es indicado cuando se emplean modelos de ecuaciones estructurales (análisis factorial confirmatorio y modelos causales entre variables latentes). Sin embargo, el correcto entendimiento de este hecho puede ayudar a los investigadores en su estrategia de análisis, reconociendo si sus indicadores son reflectivos o formativos, y cómo deben calcularse los promedios de las escalas en función de los valores de los ítems.

Así, si procedemos de forma habitual en este tipo de metodologías, asumiendo que: a) no existen más causas que provocan variación en el indicador observable; b) el valor medio del ruido blanco θ_j es cero; c) se utiliza una escala Likert arbitraria que representa a su vez una medida arbitraria de la variable latente; d) el factor de desplazamiento de escala u ordenada en el

origen α es cero; y e) que la carga factorial λ_1 del mejor indicador se fija a 1, es decir, que el indicador está en la misma escala que la variable latente, entonces retomando la ecuación [1] con solo dos indicadores (para hacerlo más sencillo), y asumiendo que los datos están recalculados como desviaciones con respecto a la media, llegamos, para el primer indicador, a lo siguiente [3]:

$$\begin{aligned} E(y_1) &= \alpha_1 + \lambda_1 E(\eta) + E(\theta_1) \\ E(\theta_1) &= 0; \alpha_1 = 0; \lambda_1 = 1 \end{aligned} \quad [3]$$

Por tanto [4]:

$$E(y_1) = E(\eta) \quad [4]$$

Para el segundo indicador [5]:

$$E(y_2) = \alpha_2 + \lambda_2 E(\eta) + E(\theta_2) \quad [5]$$

Las asunciones son [6]:

$$E(\theta_2) = 0; \alpha_2 = 0 \quad [6]$$

Ya no se fija a 1 la carga factorial, ya que de lo que se trata es de testar las restricciones causales que implica el modelo con más de un ítem, es decir, en qué medida el añadido de más ítems es válido.

Por tanto [7]:

$$E(y_2) = \lambda_2 E(\eta) \quad [7]$$

De este modo, en un modelo reflectivo de un solo factor latente, como la media de la variable latente es la misma, entonces [8]:

$$E(y_1) = \frac{E(y_2)}{\lambda_2} \quad [8]$$

Así, la media de la variable latente se puede estimar usando: $E(y_1)$ o bien: $\frac{E(y_2)}{\lambda_2}$, pero no será necesariamente: $E(y_1)/2 + E(y_2)/2$, salvo que se cumpla [8]. Para inspeccionar [8] con los valores muestrales se pueden obtener la media de y_1 ($\bar{y}_1$) y su correspondiente error estándar $\delta\bar{y}_1$, la media de y_2 ($\bar{y}_2$) y su correspondiente error estándar $\delta\bar{y}_2$ y la estimación de la carga factorial que da el programa estadístico λ_2 y su correspondiente error estándar $\delta\lambda_2$.

Se podría entonces usar la propagación de errores para comparar ambos lados de la ecuación [8], quedando [9]:

$$\bar{y}_1 \pm \delta\bar{y}_1 = \frac{\bar{y}_2}{\lambda_2} \pm \left| \frac{\bar{y}_2}{\lambda_2} \right| \sqrt{\left(\frac{\delta\bar{y}_2}{\bar{y}_2} \right)^2 + \left(\frac{\delta\lambda_2}{\lambda_2} \right)^2} \quad [9]$$

La comparación estadística de ambos lados de la ecuación [9] es una forma de testar si ambas expresiones son equivalentes, y es uno de los pasos que los

autores deben emplear antes de calcular una “media global” con los ítems de una variable latente con indicadores reflectivos.

Por tanto, es esencial que se reporten en las publicaciones las cargas factoriales (sin estandarizar), con el fin de que se pueda comprobar que, efectivamente, se cumplen las condiciones para computar una media de todos los ítems de cada variable latente.

La fijación de la varianza de error del mejor indicador, también llamado “gold standar”²⁹, implica simplemente que [10]:

$$\begin{aligned} Var[y_1] &= \lambda_1^2 Var[\eta] + Var[\theta_1] \\ Var[y_1] &= 1^2 Var[\eta] + a \end{aligned} \quad [10]$$

Donde $Var[\theta_1] = a$, y también se fija la carga factorial $\lambda_1 = 1$. El valor de a se puede tomar como un porcentaje pequeño de la varianza del indicador observable²⁹.

En las publicaciones de Andes Pediatría y Anales de Pediatría donde aparecen análisis factoriales confirmatorios^{2,4,6,8,9,10,21,23} no se hace referencia a ninguno de estos procedimientos. Las investigaciones^{2,9} muestran cargas factoriales, pero lo hacen (aparentemente) con estandarización, por lo que los lectores no pueden valorar las cargas sin estandarizar y, por ende, el procedimiento que acabamos de describir.

Validez de una escala ligada al ajuste del modelo teórico

Una vez especificados los indicadores y su relación con las variables latentes, el análisis factorial confirmatorio es habitualmente empleado como herramienta de validación empírica de un modelo teórico de ítems que representan la manifestación de variables latentes. Sin embargo, y además de la validez de contenido del instrumento de medición, se necesita que el modelo supere la prueba de la chi-cuadrado, que compara el ajuste de la matriz de covarianzas observada con la implicada por el modelo. La prueba de la chi-cuadrado es la más adecuada para detectar la mala especificación de un modelo y, si el modelo es correcto, no será rechazado incluso con tamaños de muestra más grandes²⁹⁻³⁴. Recordemos que el grado de mal ajuste en la estructura de covarianza no está asociado al grado de mala especificación³⁵, por lo que es un error identificar la divergencia estadística sobre el modelo propuesto como un indicador de grado de idoneidad de este. Además, es inadecuado utilizar solo el análisis factorial para validar una escala, sin considerar la prueba de un modelo con causas y/o efectos de esas variables latentes de la escala³⁵. Hayduk da una explicación sencilla de la relevancia de esta aproximación en contraste con las habituales formas de proceder sobre la validez de constructo, discriminante y de criterio³⁵.

Además, el análisis factorial exploratorio es incapaz de detectar la estructura real de los datos, es decir, de identificar el modelo que ha generado esos datos empíricos. Por tanto, no se debería evaluar la validez de constructo con un análisis exploratorio, donde simplemente se obtiene una visión de la estructura correlacional de los datos³⁶.

En los artículos publicados en *Andes Pediátrica* y *Anales de Pediatría*, solo el estudio de Ceniza-Bordallo y cols.² reporta correctamente un ajuste exacto del modelo. El resto no cumple con este requisito ya que, o confían en índices aproximados^{8,21,23}, los cuales son ampliamente criticados^{29-35,38}, o no reportan ningún índice de ajuste^{4,6,8,10,13}, o evalúan la validez con un análisis factorial exploratorio^{20,22,26}. Son, por tanto, modelos mal especificados que, sin embargo, se reportan como válidos, con el consiguiente sesgo en la interpretación de los resultados.

Fiabilidad usando el alfa de Cronbach

La validez es condición indispensable para un instrumento de medida. Si el instrumento es válido, es deseable que sea fiable, aunque si no es válido ya no importa la fiabilidad.

Uno de los procedimientos ampliamente empleados para valorar la fiabilidad es el alfa de Cronbach. Sin embargo, es incorrecto si el instrumento se compone de varias dimensiones o factores y, si es unidimensional, es incorrecto si no existe tau-equivalencia³⁹. Los ítems que reflejan una variable latente muestran tau-equivalencia cuando sus cargas factoriales son iguales.

Si partimos de [1], y tenemos un modelo sencillo de un factor latente con indicadores reflectivos e, entonces [11]:

$$E(y_j) = \alpha_j + \lambda_j E(\eta) + E(\theta_j) \quad [11]$$

Donde α_j es la ordenada en el origen, λ_j es la carga factorial, θ_j es ruido blanco y $E(\cdot)$ son las esperanzas matemáticas. En los modelos de covarianza, los datos de entrada para el software de ecuaciones estructurales se suelen dar en desviaciones con respecto a la media, para probar solo relaciones de covarianza, lo que hace desaparecer la ordenada en el origen al restar las ecuaciones [1] y [11], quedando [12]:

$$y_j - E(y_j) = \lambda_j [\eta - E(\eta)] + \theta_j \quad [12]$$

Tomando varianzas, y siempre que la covarianza entre el ruido blanco y la variable independiente sea nula, queda [13]:

$$Var[y_j - E(y_j)] = \lambda_j^2 Var[\eta - E(\eta)] + Var[\theta_j] \quad [13]$$

Este tipo de estructura de datos de salida es la que suelen dar los programas informáticos de ecuaciones estructurales para los modelos de medida.

En un modelo tau-equivalente, todos los λ_j son iguales, es decir, las cargas factoriales para el caso del ejemplo serían: $\lambda_1 = \lambda_2 = \lambda_3$. Solo bajo este criterio sería adecuado aplicar el alfa de Cronbach.

En los artículos publicado en *Andes Pediátrica* y *Anales de Pediatría* que emplean el alfa de Cronbach en instrumentos de medida con variables latentes, ninguno de ellos prueba la tau-equivalencia, y en varios se reporta incorrectamente el alfa de Cronbach de la escala global cuando existe multidimensionalidad^{12,6,9,11,12,18,19,21}.

Análisis factorial frente a componentes principales.

Rotación de factores

El modelo de medida mostrado en [1] se corresponde con la especificación de un análisis factorial, ya sea de naturaleza exploratoria (donde habría que añadir η_p factores diferentes en la ecuación, con sus correspondientes λ_{jp}) o confirmatoria (donde en la ecuación [1] se considera unidimensionalidad). La interpretación causal es pertinente, ya que es la variación en la variable latente η la que produce variación en los indicadores observables y_j .

Sin embargo, la estructura de un análisis de componentes principales es diferente, tal como se observa en [14]:

$$\eta_j^* = \sum_{p=1}^j \lambda_{jp}^* y_p \quad [14]$$

Ahora se tienen η_j^* que son una combinación lineal de y_p variables observables, con la condición de que esos factores estén no correlacionados entre ellos, aunque se pueden especificar rotaciones que no son ortogonales, sino oblicuas, lo que permite la correlación. El asterisco significa que ni los factores ni las cargas factoriales tienen por qué coincidir con [1].

Rápidamente se observa que el sentido causal cambia por completo con respecto a [1], siendo similar a la especificación formativa detallada en [2]. Ahora son las variables observables las que provocan variación en los factores latentes (llamados componentes principales). Por tanto, la implementación de análisis de componentes principales es incompatible con una visión clásica de la medición, en el que el instrumento de medida es sensible a la variación de lo que se quiere medir (la variable latente). De este modo, es inadecuado para una especificación reflectiva de los indicadores.

Además, la consideración de ortogonalidad de esos componentes principales es una restricción extremadamente fuerte para las estructuras factoriales multidimensionales, en las que es muy habitual que los factores tengan cierto grado de correlación^{40,41}. Por tanto, a no ser que exista una base teórica sólida detrás, las

rotaciones de componentes principales deberían ser oblicuas. Evidentemente, la consideración de componentes principales es incompatible con el uso del alfa de Cronbach⁴².

En *Andes Pediatría* y *Anales de Pediatría* se han publicado artículos aplicando al mismo tiempo el alfa de Cronbach y componentes principales^{3,11,24,25}, relacionando validez de constructo con componentes principales¹¹, aplicando componentes principales cuando la especificación causal debiera ser [1] y no [14]¹⁴, y realizando inadecuadas rotaciones ortogonales^{11,24}.

Discusión

Cuando se emplean escalas de medida, un adecuado procedimiento de modelización y análisis es fundamental para la validez de los resultados. Este trabajo ha mostrado que se necesita ser más preciso y riguroso tanto en la conceptualización como en los análisis de este tipo de instrumentos de medida en las publicaciones de *Andes Pediatría* y *Anales de Pediatría*.

Este trabajo ha analizado el nivel de cumplimiento de determinados aspectos metodológicos y estadísticos en los estudios de validación de instrumentos psicométricos en el ámbito de la pediatría, con ánimo totalmente constructivo. Es así como se han expuesto algunas ecuaciones básicas sobre las que se fundamentan las decisiones metodológicas en esos estudios, siendo, por tanto, útiles tanto para autores como para revisores.

Hay que reconocer, asimismo, que en psicometría existen acalorados debates sobre visiones contrapuestas relacionadas, y algunos de los tópicos comentados en este trabajo han sido objeto de ello^{27,31,32}. No obstante, se recomienda encarecidamente a los autores que visionen las referencias adjuntas y respondan con una visión crítica, admitiendo también que este estudio abarca solo una parte de las diferentes formas de evaluación de las propiedades de un instrumento psicométrico. A este respecto, el consenso COSMIN (Consensus-based Standards for the selection of health Measurement INstruments) es una iniciativa internacional que tiene como objetivo mejorar la selección de instrumentos de medición de resultados de

salud, agrupa estas evaluaciones en cuatro dominios: validez, confiabilidad, capacidad de ser respondido e interpretabilidad⁴³. De este modo, este trabajo se centra en aspectos integrados en los dos primeros.

De gran importancia en la evaluación de cualquier instrumento de medida, una vez determinada la correcta relación entre los indicadores observables y las variables latentes, es el análisis de la validez testada con índices de ajuste exacto. Este es un tema que la gran mayoría de los autores sistemáticamente obvian, quizá porque es más “cómodo” confiar en índices de ajuste aproximado para así aceptar los modelos que proponen que, de otro modo, quedarían rechazados (y por tanto no se podrían realizar ningún análisis posterior sobre medias de indicadores, análisis de grupos, etc.). Sin embargo, los argumentos de las investigaciones^{35-38,44} son lo suficientemente sólidos como para cambiar la tendencia de publicar modelos que fallan al testarse contra los datos empíricos, cuestionando de forma importante los resultados e interpretaciones posteriores.

Así, “confiar” en instrumentos supuestamente validados para, por ejemplo, adaptarlos al contexto español tampoco es una estrategia potencialmente libre de sesgo. En primer lugar, habría que tener una visión crítica de cómo se han validado esos modelos originales, ya que, si su estrategia de validación falla en alguno de los puntos considerados en este trabajo, es probable que esos instrumentos no sean válidos, aunque se hayan publicado como tales. Y, en segundo lugar, habría que proceder a su validación en el nuevo contexto de estudio. Así, prestar atención a los elementos brevemente ilustrados en este estudio con los contraejemplos publicados en *Andes Pediatría* y *Anales de Pediatría*, debiera ser una forma de avanzar en la investigación pediátrica basada en cuestionarios. La rigurosidad en este tipo de procedimientos es esencial para ofrecer interpretaciones correctas que deriven en mejores intervenciones.

Conflicto de intereses

Los autores declaran no tener conflicto de intereses.

Referencias

- Bollen KA. Structural Equations with Latent Variables. Wiley Series in Probability and Mathematical Statistics. New York: Wiley 1989;1-514. <https://doi.org/10.1002/9781118619179>
- Ceniza-Bordallo G, Gómez Fraile A, Martín-Casas P, et al. Validation of the Spanish version of the Pain Catastrophizing Scale for Children (PCS-C). *An Pediatr (Barc)*. 2023;99(5):295-303. <https://doi.org/10.1016/j.anpedi.2023.09.011>
- Balaguer-Martínez JV, Esquivel-Ojeda JN, Valcarce-Pérez I, et al. Traducción al español y validación de una escala para la observación de una toma de lactancia materna: la «Bristol Breastfeeding Assessment Tool». *An Pediatr (Barc)*. 2021;S1695-4033(21)00181-8. Advance online publication. <https://doi.org/10.1016/j.anpedi.2021.04.012>
- Godoy-Molina E, Fernández-Ferrández T, Ruiz-Sánchez JM, et al. A scale for the identification of the complex chronic pediatric patient (PedCom Scale): A pilot study. *An Pediatr (Barc)*. 2022;97(3):155-60. <https://doi.org/10.1016/j.anpedi.2021.10.004>
- Jiménez Candel MI, Carpena Lucas PJ, Ceballos-Santamaría G, et al. Relationship between modifiable risk factors and overweight in adolescents aged 12-14 years. *An Pediatr (Barc)*. 2021;95(3):159-66. <https://doi.org/10.1016/j.anpedi.2020.08.008>
- Nievas Soriano BJ, García Duarte S, Fernández Alonso AM, et al. Coste y ahorro potencial generado por una web de eSalud de pediatría para padres. *An Pediatr (Barc)*. 2021;94(2):92-8. <https://doi.org/10.1016/j.anpedi.2020.03.014>
- Mardomingo Sanz MJ, Sancho Mateo C, Soler López B, en representación del grupo de investigación del estudio SELFIE, & Anexo 1. Grupo de estudio SELFIE. Evaluación de la comorbilidad y la ansiedad social en adolescentes con trastorno por déficit de atención con hiperactividad: Estudio SELFIE. *An Pediatr (Barc)*. 2019;90(6):349-61. <https://doi.org/10.1016/j.anpedi.2018.07.014>
- López-Villalobos JA, Sacristán-Martín AM, Garrido-Redondo M, et al. Calidad de vida relacionada con la salud en casos de trastorno por déficit de atención con hiperactividad con/sin tratamiento farmacológico. *An Pediatr (Barc)*. 2019;90(5):272-9. <https://doi.org/10.1016/j.anpedi.2018.04.016>
- Orgilés M, Melero S, Penosa P, et al. Calidad de vida relacionada con la salud informada por los padres en preescolares españoles: propiedades psicométricas del Kiddy-KINDL-R. *An Pediatr (Barc)*. 2019;90(5):263-71. <https://doi.org/10.1016/j.anpedi.2018.04.019>
- Martín Martín R, Sánchez Bayle M, Teruel de Francisco C. Aspectos psicosociales relacionados con la hiperfrecuentación en las consultas pediátricas de atención primaria. *An Pediatr (Barc)*. 2019;90(1):26-31. <https://doi.org/10.1016/j.anpedi.2018.03.008>
- Gironés Muriel A, Campos Segovia A, Ríos Gómez P. Estudio de validación y fiabilidad del cuestionario de preocupación paterna sobre la cirugía. ¿Qué preocupa a los padres? *An Pediatr (Barc)*. 2018;88(1):24-31. <https://doi.org/10.1016/j.anpedi.2017.01.007>
- Gascón-Cánovas JJ, Russo de Leon JR, Cózar Fernández A, et al. Adaptación cultural al español y baremación del Adolescent Peer Relations Instrument (APRI) para la detección de la victimización por acoso escolar: Estudio preliminar de las propiedades psicométricas. *An Pediatr (Barc)*. 2017;87(1):9-17. <https://doi.org/10.1016/j.anpedi.2015.12.003>
- Martín Martín R, Sánchez Bayle M, Gancedo García C, et al. Las familias de la crisis en las consultas pediátricas de Atención Primaria: estudio descriptivo observacional. *An Pediatr (Barc)*. 2016;84(4):189-94. <https://doi.org/10.1016/j.anpedi.2015.06.005>
- Ullan AM, Perelló M, Jerez C, et al. Validación de la versión española de la escala de evaluación del dolor postoperatorio Parent's Postoperative Pain Management. *An Pediatr (Barc)*. 2016;84(2):106-13. <https://doi.org/10.1016/j.anpedi.2015.04.001>
- Ordóñez A, Maganto C, González R. Quejas somáticas, conciencia emocional e inadaptación en población escolar. *An Pediatr (Barc)*. 2015;82(5):308-15. <https://doi.org/10.1016/j.anpedi.2014.03.020>
- Ramos Fernández JM, Cerdón Martínez A, Galindo Zavala R, et al. Validación de una escala clínica de severidad de la bronquiolitis aguda. *An Pediatr (Barc)*. 2014;81(1):3-8. <https://doi.org/10.1016/j.anpedi.2013.06.020>
- Magallón S, Crespo-Eguílaz N, Ecay M, et al. Estilo comportamental al inicio del segundo año de vida: estudio retrospectivo en escolares afectados de trastorno por déficit de atención e hiperactividad. *An Pediatr (Barc)*. 2009;70(6):562-9. <https://doi.org/10.1016/j.anpedi.2009.01.018>
- Escobar R, Lucero N, Solares C, et al. Escala de evaluación funcional de extremidades superiores en niños con distrofia muscular de Duchenne y Atrofia muscular espinal. *Rev Chil Pediatr*. 2017;88(1):92-9.
- Andrade RD, Bustos NC, Brito CH, et al. Evaluación psicométrica de la escala de autoeficacia de la lactancia materna. *Andes pediátr*. 2022;93(4):470-76. <https://dx.doi.org/10.32641/andespediatr.v93i4.3474>
- Navarro-Tapia S, Ramirez M, Claverias C, et al. Validación de "The Parental Stressor Scale Infant Hospitalization modificado, al español" en una unidad cuidados intensivos pediátricos. *Rev Chil Pediatr*. 2019;90(4):399-10. <https://dx.doi.org/10.32641/rchped.v90i4.1020>
- Cabezas TA, Rivera LC, Sarmiento LL. Validación psicométrica de Pauta de Clasificación de Complejidad Médica de niños, niñas y adolescentes con Necesidades Especiales de Atención en Salud. *Andes pediátr*. 2023;94(5):638-45. <https://dx.doi.org/10.32641/andespediatr.v94i5.4522>
- Funes DF, León LF, Valenzuela CR. Evaluación de conocimientos y autonomía para la transición de adolescentes hacia la atención de adultos. *Rev Chil Pediatr*. 2020;91(5):722-31. <https://dx.doi.org/10.32641/rchped.v91i5.1519>
- Alvarez C, Briceño A, Álvarez K, et al. Estudio de adaptación y validación transcultural de una escala de satisfacción con la vida para adolescentes. *Rev Chil Pediatr*. 2018;89(1):51-8. Recuperado de: <https://www.revistachilenadepediatria.cl/index.php/rchped/article/view/198> 22 dic. 2023
- Morán KJ, Lecannelier AF, Rodríguez TJ. Estudio de validación de un cuestionario de apego en adolescentes. *Rev Chil Pediatr*. 2014;85(4):437-42. Recuperado de: <https://www.revistachilenadepediatria.cl/index.php/rchped/article/view/3142> 22 dic. 2023
- Morales MP, Santos MJ, González SA, et al. Validación factorial de un cuestionario para medir la conducta de comer en ausencia de hambre y su asociación con obesidad infantil. *Rev Chil Pediatr*. 2012;83(5):431-37. Recuperado de: <https://www.revistachilenadepediatria.cl/index.php/rchped/article/view/2886> 22 dic. 2023
- Olivari C, Guzmán-González M. Validación del cuestionario general de búsqueda de ayuda para problemas de salud mental en adolescentes. *Rev Chil Pediatr*. 2017;88(3):324-31. Recuperado de: <https://www.revistachilenadepediatria.cl/index.php/rchped/article/view/292> 22 dic. 2023
- Bollen KA, Diamantopoulos A. In defense of causal-formative indicators: A minority report. *Psychol Methods*. 2017;22(3):581-96. <https://doi.org/10.1037/met0000056>
- Edwards JR, Bagozzi RP. On the nature and direction of relationships between constructs and measures. *Psychol*

- Methods. 2000;5(2):155-74. <https://doi.org/10.1037/1082-989X.5.2.155>
29. Hayduk LA. LISREL Issues, Debates, and Strategies. Baltimore: Johns Hopkins University Press; 1996;25-30. <https://doi.org/10.56021/9780801853364>
 30. Hayduk LA, Littvay L. Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? BMC Med Res Method. 2012;12(159):1-17. <https://doi.org/10.1186/1471-2288-12-159>
 31. Hayduk LA, Glaser DN. Jiving the four-step, waltzing around factor analysis, and other serious fun. Struct Equ Modeling. 2000;7(1):1-35. https://doi.org/10.1207/S15328007SEM0701_01
 32. Hayduk LA, Glaser DN. Doing the four-step, right-2-3, wrong-2-3: A brief reply to Mulaik and Millsap; Bollen; Bentler; and Herten and Costner. Struct Equ Modeling, 2000;7(1):111-23. https://doi.org/10.1207/S15328007SEM0701_06
 33. Hayduk LA, Cummings G, Boadu K, et al. Testing! testing! one, two, three - Testing the theory in structural equation models! Pers Individ Dif, 2007;42(5):841-50. <https://doi.org/10.1016/j.paid.2006.10.001>
 34. Antonakis J, Bendahan S, Jacquart P, et al. On making causal claims: A review and recommendations. Leadersh Q. 2010;21:1086-20. <https://doi.org/10.1016/j.leaqua.2010.10.010>
 35. Hayduk LA. Shame for disrespecting evidence: the personal consequences of insufficient respect for structural equation model testing. BMC Med Res Methodol. 2014;14:124. <https://doi.org/10.1186/1471-2288-14-124>
 36. Hayduk LA. Seeing perfectly-fitting factor models that are causally misspecified: Understanding that close-fitting models can be worse. Educ Psychol Meas. 2014;74(6):905-26. <https://doi.org/10.1177/0013164414527449>
 37. Hayduk LA, Glaser DN. Jiving the four-step, waltzing around factor analysis, and other serious fun. Struct Equ Modeling. 2000;7:1-35. https://doi.org/10.1207/S15328007SEM0701_01
 38. Barret P. Structural equation modelling: Adjudging model fit. Pers Individ Dif. 2007;42(5):815-24. <https://doi.org/10.1016/j.paid.2006.09.018>
 39. Cho E. Making reliability reliable: A systematic approach to reliability coefficients. Organ Res Method. 2016;19:651-82. <https://doi.org/10.1177/1094428116656239>
 40. Thurstone LL. Multiple factor analysis. 1947. Chicago: University of Chicago Press.
 41. Browne MW. An overview of analytic rotation in exploratory factor analysis. Multivariate Behav Res. 2001;36:111-50. https://doi.org/10.1207/S15327906MBR3601_05
 42. Bollen KA, Lennox R. Conventional wisdom on measurement: A structural equation perspective. Psychol Bull. 1991;110(2):305-14. <https://doi.org/10.1037/0033-2909.110.2.305>
 43. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45. <https://doi.org/10.1016/j.jclinepi.2010.02.006>
 44. Martínez JA, Martínez L. El análisis factorial confirmatorio y la validez de escalas en modelos causales. An Psicol 2009;25(2):368-74. Recuperado a partir de <https://revistas.um.es/analesps/article/view/88081>