

**Validez de contenido usando la V de Aiken con intervalos de confianza: aportes a Rodríguez et al.**

*Content validity using Aiken's V with confidence intervals: contributions to Rodríguez et al.*

**Sr. Editor:**

Recientemente, con el objetivo de brindar alternativas ante la escasez de autoinformes para evaluar la actividad física, Rodríguez et al.<sup>1</sup> realizaron la validación del contenido del *Previous Day Physical Activity Recall* (PDPAR) mediante la valoración de sus aspectos semánticos, idiomáticos y contextuales de su traducción por parte de jueces expertos, y si bien reportan un alto grado de acuerdo entre ellos, el índice de validez de contenido (ICV) empleado carece de la rigurosidad necesaria para considerarlo como una medida para determinar si existen o no evidencias de validez de contenido, lo que puede poner en duda las conclusiones de los autores respecto a la calidad de los ítems. En tal sentido, se propone que una alternativa con mayor sustento teórico y empírico es la V de Aiken<sup>2</sup>, pero utilizando además intervalos de confianza (IC).

Brevemente, la V busca cuantificar el acuerdo entre jueces considerando como un indicador importante el promedio de calificaciones ( $\bar{X}$ ), pero toma en cuenta además la calificación mínima posible ( $l$ ) y el rango de calificaciones posibles ( $k = \text{máx} - \text{mín}$ ):

$$V = \frac{\bar{X} - l}{k}$$

El valor de la V puede oscilar entre cero y la unidad, y mientras más próximo se encuentre a esta, se infiere que el acuerdo entre los jueces es mayor, y por tanto, las evidencias de validez de contenido serán mayores. No obstante, como cualquier estadístico, la V está influida por el error muestral, por lo cual es necesario establecer el rango de posibles valores que asumiría el parámetro. Ante esa posibilidad, los IC son una aproximación más útil. Existe un método para construir intervalos de confianza asimétricos para la V,<sup>3</sup> y con esos resultados podría valorarse de forma más precisa el acuerdo entre los jueces.

$$L = \frac{2nkV + z^2 - z\sqrt{4nkV(1-V) + z^2}}{2(nk + z^2)}$$

$$U = \frac{2nkV + z^2 + z\sqrt{4nkV(1-V) + z^2}}{2(nk + z^2)}$$

Donde L y U son los límites inferior y superior del IC para la V, Z es el valor en distribución estándar correspondiente al 90%, 95% o 99%, y n es el número de jueces. Si bien no hay un consenso respecto a qué valor del límite inferior del IC de la V puede ser considerado como aceptable, la evidencia parece indicar que 0,7 sería adecuado,<sup>4</sup> ya que no permite que sea valorado como *representativo* un ítem cuando el acuerdo entre jueces no es lo suficientemente alto. Por ejemplo, si una V es de 0,72, pero su límite inferior es 0,63, se consideraría que ese ítem presenta un contenido poco representativo.

Para realizar el análisis complementario de la V con IC a través de un módulo especializado<sup>5</sup> se cuentan con los datos necesarios: cantidad de jueces (siete), los promedios de calificación para cada ítem (tabla 2, p. 201), y el rango de calificación, que va de *no representativo* (1) hasta *representativo* (4). Por razones de economía de espacio, solo se calculará la V con IC para los valores promedio más frecuentes, y el lector deberá hacer la correspondencia con cada uno de los ítems de la *Tabla 2* del manuscrito<sup>1</sup>.

TABLA 1. Cálculo de las V de Aiken con intervalos de confianza

Valor promedio de la calificación	V	IC 95% V
3,14	0,713	<b>0,499 - 0,861</b>
3,43	0,810	<b>0,601 - 0,924</b>
3,57	0,857	<b>0,653 - 0,950</b>
3,71	0,903	0,709 - 0,973
3,86	0,953	0,775 - 0,992
4,00	1,000	0,845 - 1,000

Según los resultados, los V correspondientes a los promedios 3.14, 3.43, y 3.57 presentan IC que caen por debajo de lo esperado (.70), lo cual indica que si bien el acuerdo entre jueces fue aparentemente bueno, al analizarlo mediante este procedimiento se concluye que dichos ítems necesitarían una revisión adicional. De este modo, los ítems 1, el *Guión de instrucciones* y la *Escala de actividades* presentan indicadores al menos cuestionables en cuanto a su *Equivalencia semántica-idiomática*; y los ítems 24, 26 y 28, en lo que respecta a su *Equivalencia cultural*.

Si bien el contenido de esta carta pueda alejarse temáticamente del foco de la revista, ante el incremento de estudios de validación que comienzan con la valoración del contenido,

estos métodos pueden ser de utilidad ya que brindan resultados más precisos en esta primera etapa, ya que si no se valora de forma apropiada el contenido de los ítems, cualquier proceso posterior se verá afectado, inclusive análisis estadísticos tales como comparación de grupos, correlaciones, etc, y posteriormente la toma de decisiones acerca de la persona evaluada.

Mg. Sergio Alexis Domínguez-Lara  
Instituto de Investigación de Psicología,  
Universidad de San Martín de Porres,  
Lima - Perú.

1. Rodríguez PL, Pérez JJ, García E, Rosa A. Adaptación transcultural de un cuestionario que evalúa la actividad física en niños de 10 y 11 años. *Arch Argent Pediatr* 2015;113(3): 198-204.
2. Aiken LR. Three coefficients for analyzing the reliability and validity of ratings. *Educ Psychol Meas* 1985; 45(1):131-42.
3. Penfield R, Giacobbi Jr. P. Applying a score confidence interval to Aiken's item content-relevance index. *Meas Phys Educ Exerc Sci* 2004; 8(4): 213-25.
4. Charter RA. A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. *J Gen Psychol* 2003; 130(3): 290-304.
5. Merino Soto C, Livia Segovia J. Intervalos de confianza asimétricos para el índice la validez de contenido: Un programa Visual Basic para la V de Aiken. *An Psicol* 2009; 25(1): 169-71.

En respuesta

A la atención de D. Mg. Sergio Alexis Domínguez-Lara,

Estimado investigador:

En primer lugar, le agradezco las atenciones que nos ha dispensado con su crítica metodológica al artículo publicado en la revista. Es muy valioso y constructivo que la actividad investigadora se vea sometida a evaluación y valoración por parte de investigadores ajenos a los trabajos que, sin duda, aportan contribuciones que inciden en la mejora progresiva y en la evolución constante de la calidad científica.

Sin embargo, es importante que, todo juicio de valor realizado sobre trabajos de investigación de compañeros ajenos a nuestra influencia se centre, en primer lugar, en destacar los aspectos positivos y agradecer la contribución que se hace, aunque sea modesta, al ámbito del conocimiento de los constructos analizados y estudiados. Usted se lanza de manera directa a lo que, bajo su punto de vista, es deficitario en el trabajo. En este sentido, en relación a los argumentos metodológicos

que usted aporta, aunque los pueda rebatir, son aportaciones que mejoran mi sustento formativo en materia de investigación.

En lo que respecta al artículo, mencionar que el *Previous Day Physical Activity Recall* (PDPAR) ya ha sido validado inicialmente por los autores que la construyeron<sup>1,2</sup> y, posteriormente, hay en la literatura varios trabajos que la adaptan a distintos contextos. Al ser una escala de gran utilidad y versatilidad en el contexto escolar, decidimos llevar a cabo un complejo proceso de adaptación a la cultura española, envolviendo a 7 expertos, y siguiendo las directrices de Beaton et al. (2000)<sup>3</sup>. Se decidió emplear el índice de validez de contenido (IVC), para observar la proporción de grado de acuerdo entre jueces respecto a la validez semántica, idiomática y contextual, en base al estudio de revisión que realizó Alexandre y Coluci (2011), donde se puede observar que es un indicador adecuado para dicha causa;<sup>4</sup> además existen otras publicaciones similares que ya lo han implementado en la literatura científica.<sup>5</sup>

Como se indica en el artículo, la traducción de los 33 apartados de la lista de actividades, más el guión de aplicación, la escala de actividades y la tabla de registro, se llevó a cabo de forma satisfactoria. En prácticamente todos los apartados, la media de las respuestas de los 7 exploradores, valorando de 1 a 4, fue de 4 puntos; y se tuvo en cuenta el criterio de eliminación o revisión si el promedio bajaba de 3 puntos.<sup>4</sup> Cuando se aplicó el IVC, la mayoría de los ítems tanto para su equivalencia semántica/idiomática como cultural mostraron valores de 1.0.

Como bien indica, en 3 casos de los 36 analizados, aplicando de forma rigurosa la V de Aiken con intervalos de confianza (IC), los resultados del IC parecen estar por debajo de los esperados, y podrían ser objeto de una revisión adicional, pese a que el promedio de valoración por parte de los 7 jueces fuera de 3.14, 3.43, y 3.57 (sobre 4). En cualquier caso, es preciso recordar, que tras el proceso de adaptación transcultural se realizó un pre-test con 32 escolares, para observar si los mismos entendían el cuestionario, obteniendo resultados satisfactorios, en cuanto a su comprensión idiomática y cultural, en todos los casos; tan sólo 4 mostraron dificultades con otros procesos ajenos a la comprensión del texto.

Si bien la V de Aiken se presenta como un proceso más dentro de los posibles para la validez del contenido<sup>6</sup>, parece ser interesante para futuras validaciones que se realicen en torno a este tipo de

cuestionarios y será tenida en cuenta, pese a que no se observe un criterio unificado en la literatura internacional en torno a qué prueba utilizar para este tipo de análisis.

*Prof. Pedro Luis Rodríguez García*  
Facultad de Educación.

Universidad de Murcia (España).

Director del Grupo de Investigación AFYDOS.

1. Weston AT, Petosa R, Pate R. Validation of an instrument for measurement of physical activity in youth. *Med Sci Sports Exerc* 1997;29(1):138-43.
2. Trost SG, Ward DS, McGraw B, Pate R. Validity of the Previous Day Physical Activity Recall (PDPAR) in Fifth-Grade Children. *Pediatr Exerc Sci* 1999;11(4):341-8.
3. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000;25(24):3186-91.
4. Alexandre NMC, Coluci MZO. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. *Cien Saude Colet* 2011;16(7):3061-8.
5. São-João TM, Rodrigues RC, Gallani MC, Miura CT, et al. Cultural adaptation of the Brazilian version of the Godin-Shephard Leisure-Time Physical Activity Questionnaire. *Rev Saude Publica* 2013;47(3):479-87.
6. Pedrosa, I., Suárez-Álvarez, J. y García-Cueto, E. Evidencias sobre la validez de contenido: avances teóricos y métodos para su estimación. *Acción Psicol*, 2014;10(2):3-20.