Back to content-based validity

De regreso a la validez basada en el contenido

José Ventura-León*.

*Universidad Privada del Norte, Lima, Perú.

The development and validation of self-reported measuring instruments as a way of quantifying addictive behaviors is currently common practice. These instruments in articles about addiction, however, are dominated by methods involving structural equations and, therefore, evidence based on internal structure (Mezquita, Camacho, Suso-Ribera, Ortet & Ibáñez, 2018; Mezquita, Ruiz-Valero, Martínez-Gómez, Ibáñez & Ortet-Fabregat, 2019). This is the case even though it is known that there are other sources of evidence, such as test consequence, relationships with other variables, content and response processes, which are considered to contribute methodological rigor when researching addiction (Fonseca, 2017). The purpose of this letter to the editor is thus to examine the importance of content-based validity in the development and/or adaptation of measurement instruments in the study of addiction.

Suppose you wanted to create a scale to measure addiction to love and an item is worded: "I experience anguish when my partner is not with me", with a factorial load of greater than .30; despite this, the item may not capture a behavior representative of the universe of behaviors of the construct in question (Cohen & Swerdik, 2001). To test this, the researcher needs the judgment and assessment of an expert, someone who can be considered as such because of his or her extensive experience and recognition in the field (Escobar-Pérez & Cuervo-Martínez, 2008).

The procedure by which the logic is reviewed or by which the representativeness and relevance of the test contents in the interpretation of the test scores is analyzed, is called *content-based validity* (American Psychological Association [APA], American Educational Research Association [AERA], and National Council on Measurement in Education [NCME], 2014). Such a review of the representativeness or relevance of the construct can prevent the covariation of erroneous theoretical information (Haynes, Richard & Kubany, 1995), thus avoiding irrelevant construct variance (APA, AERA & NCME, 2014), which is important because measurement instruments should rely not only on factor models, but also on theoretical argumentation which can show whether an item is representative of a particular domain or not (Bonifay, Lane & Reise, 2017).

Given the above, the author of this letter would like to offer readers an expert rater grid, which can be requested in its entirety and free of charge (see an excerpt in Appendix A). This grid is based on APA, AERA and NCME (2014) guidelines concerning the relevance (degree to which the item is important and should be included in the measurement of the construct) and representativeness (degree to which the item represents the construct to be measured); in addition, a clarity criterion is incorporated (the degree to which the item is clear and understandable).

Once the raters' answers have been obtained, they can be quantified by Aiken's *V*, a coefficient which is simple to calculate and easy to interpret, as expressed below (Penfield & Giacobbi, 2004):

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

Received: July 2018; Accepted: November 2018.

Send correspondence to:
Jose Ventura-León. Av. Tingo María 1122, Breña, Lima, Telf. (01)6044700 anexo: 3462.
Email: jose.ventura@upn.pe

Where $\bar{X}$ is the mean expert rating, l is the lowest possible score and k is the difference between the highest and lowest score on the rating scale. Values of V close to 1 indicate perfect agreement between the raters. A minimum cut-off point of .70 is required (Napitupulu, Syafrullah, Rahim, Amar & Sucahyo, 2018). Likewise, at present, confidence intervals (CI) can be established for Aiken's V, the mathematical expression of which is presented below (Penfield & Giacobbi, 2004):

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

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

Where L is the lower limit and U the upper limit, n is the number of raters, k is the difference between the highest and lowest scores on the scale; V is the value of Aiken's V; and z is the standard distribution chosen, so 90%, 95% and 99% confidence corresponds to it 1.65, 1.96 and 2.58 respectively.

For the interpretation of CIs, it is recommended that the value of the lower limit ≥ .70 (Charter, 2003), although it is known that CI size depends to a large extent on the increase of sample size (Penfield & Giacobbi, 2004). If you would like to calculate Aiken's V with its respective CIs, you can request an Excel® spreadsheet at no cost from the author of this letter or use the following codes in the statistical program R:

```
####AIKEN'S V#####
x = 1.90 # arithmetic mean of expert ratings
l = 0 # lowest
s = 3 # highest value
k = s-l # range
v = (x-l)/k ## Aiken's v equation
v
####CONFIDENCE INTERVALS#####
Z = 1.96 # value of z at 95%
N = 10 # number of raters
IC1 = (((2*v*N*K)+(Z^2)))
IC2 = Z*(sqrt((4*N*K*v)*(1-v)+(Z^2)))
IC3 = 2*((N*K)+(Z^2))
INFIC = (IC1- IC2)/IC3
SUPIC = (IC1+IC2)/IC3
INFIC
SUPIC
```

In conclusion, the incorporation of evidence based on content in self-reported addiction instruments is relevant for two reasons: (a) the review of item content by expert raters prior to carrying out the statistical analyses will allow construct irrelevant variance to be reduced; (b) reporting that a test is valid merely because it shows evidence of validity based on its internal structure is insufficient; it is necessary to explore more sources of validity, one of them being based on content. Finally, it is hoped that the R codes and the expert rater grid can offer a way of returning to a review of item content and coherence with theoretical postulates, thereby providing better-calibrated scales, questionnaires or tests in addiction research.

Funding

Universidad Privada del Norte.

Conflict of interests

None.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bonifay, W., Lane, S. P. & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science, 5*, 184-186. doi:10.1177/2167702616657069.

Charter, R. A. (2003). A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. *Journal of General Psychology, 130*, 290-304. doi:10.1080/00221300309601160.

Cohen, R. & Swerdlik, M. (2001). *Pruebas y evaluación psicológicas: Introducción a las pruebas y a la medición.* (4ª ed.). México: Mc Graw Hill.

Escobar-Pérez, J. & Cuervo-Martínez, A. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición, 6*, 27-36.

Fonseca, E. (2017). Methodological rigour in the study of addictions. *Adicciones, 29*, 147-149. doi:10.20882/adicciones.994.

Haynes, S. N., Richard, D. & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*, 238-247. doi:10.1037/1040-3590.7.3.238.

Mezquita, L., Camacho, L., Suso-Ribera, C., Ortet, G. & Ibáñez, M. I. (2018). Desarrollo y validación de la versión corta del cuestionario sobre expectativas de los efectos del alcohol (EQ-SF). *Adicciones, 30*, 271-281. doi:10.20882/adicciones.920.

Mezquita, L., Ruiz-Valero, L., Martínez-Gómez, N., Ibáñez, M. I. & Ortet Fabregat, G. (2019). Desarrollo y validación

de la versión breve del cuestionario de motivos de consumo de marihuana (MMM SF). *Adicciones, 31*, 106-116. doi:10.20882/adicciones.979.

Napitupulu, D., Syafrullah, M., Rahim, R., Amar, A. & Sucahyo, Y. G. (2018). Content validity of critical success factors for e-Government implementation in Indonesia. In *IOP Conference Series: Materials Science and Engineering,* 352, 1-10. doi:10.1088/1757-899X/352/1/012058.

Penfield, R. D. & Giacobbi, Jr, P. R. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in Physical Education and Exercise Science, 8*, 213-225. doi:10.1207/s15327841mpee0804_3.

Appendix A

Table 1. *Excerpt from the expert rater grid.*

		Relevance				Representativeness				Clarity				Suggestions
Positive self-esteem* Positive perception of oneself, taking into account one's qualities														
Nº	**Items**													
1	I feel I'm worth as much as others	0	1	2	3	0	1	2	3	0	1	2	3	
3	I think I have some good qualities	0	1	2	3	0	1	2	3	0	1	2	3	
4	I can do things just as well as others can	0	1	2	3	0	1	2	3	0	1	2	3	
6	I have a positive attitude towards myself	0	1	2	3	0	1	2	3	0	1	2	3	
7	I almost always feel good about myself	0	1	2	3	0	1	2	3	0	1	2	3	
Negative self-esteem Negative perception of oneself, tending to see one's bad sides														
Nº	**Items**													
2	I almost always feel like a failure	0	1	2	3	0	1	2	3	0	1	2	3	
5	I feel I haven't got much to be proud of	0	1	2	3	0	1	2	3	0	1	2	3	
8	I would like to have more self-respect	0	1	2	3	0	1	2	3	0	1	2	3	
9	I feel really useless sometimes	0	1	2	3	0	1	2	3	0	1	2	3	
10	I sometimes feel I'm no good for anything	0	1	2	3	0	1	2	3	0	1	2	3	

Note: The response options of the scale range from 1 to 4 as follows: (0) Completely disagree; (1) Disagree; (2) Agree; (3) Completely agree; *: Excerpt from the expert rater grid based on the Rosenberg scale in Spanish by Atienza, Moreno and Balaguer, 2000.

Signature of expert rater

.....