EDUCATIONIS MOMENTUM
vol 3, n.° 1, 2017, pp. 153-155. ISSN (impr.): 2414-1364; (online): 2517-9853
https://doi.org/10.36901/em.v3i1.116

Sobre la importancia de los instrumentos de evaluación en diseños preexperimentales

On the Importance of Measurement Instruments in Pre-Experimental Designs

Sergio Alexis Dominguez-Lara
Universidad de San Martín de Porres
sdominguezl@usmp.pe

Recibido: 2017.05.26
Aceptado: 2017.08.28

Estimado editor:

Al respecto del trabajo publicado por Sánchez Sanssoni (2015) en el primer volumen de su revista, me gustaría compartir algunas consideraciones de corte metodológico.

Al implementar estrategias manipulativas (Ato, López, & Benavente, 2013), sobre todo aquellas más propensas al error de muestreo (y a otras fuentes de error), como el diseño preexperimental, es recomendable poner atención a los detalles, más aún cuando el objetivo es brindar evidencias, sobre la conducta humana, a favor de un determinado modelo emergente como la pedagogía del amor y la ternura.

Con base en los resultados del citado trabajo, la autora concluye que «la aplicación del taller ha mejorado la disciplina escolar de los alumnos del grupo experimental» (Sánchez Sanssoni, 2015, p. 78). Si llegan a aceptarse esas conclusiones como válidas, ese enfoque podría ser adoptado en diversas instituciones educativas debido a que existiría evidencia de su influencia sobre la disciplina infantil. En este sentido, es conveniente destacar un aspecto crucial en el trabajo que probablemente haya sido pasado por alto: las propiedades psicométricas del instrumento de evaluación usado en la medición pre— y postintervención.

Sánchez Sanssoni (2015) reportó que mediante un instrumento *ad hoc*

evaluó orden, respeto y responsabilidad (p. 70). Sin embargo, no se describe detalladamente la prueba ni se presenta un reporte de confiabilidad y validez que garantice que cualquier inferencia derivada de las puntuaciones se refiera al constructo que se desee evaluar. En otros espacios se ha defendido esta práctica y lo esclarecedora que puede ser (Calderón-De la Cruz, Dominguez-Lara, & Navarro-Loli, 2016).

Quizá pueda argumentarse que se trata de una guía de observación, pero no por ello deja de evaluar aspectos conductuales que necesitan ser refrendados con criterios empíricos. Por ejemplo, es posible evaluar la validez de contenido (¿las conductas listadas en la guía representan conceptualmente la dimensión conductual que pretenden evaluar?) (véase Sireci & Faulkner-Bond, 2014) utilizando estadísticos como la V de Aiken (Dominguez-Lara, 2016a); o analizar la confiabilidad intercalificadores (Merino-Soto, Calderón-De la Cruz, & Manzanares, 2016), lo que permite concluir con cierto grado de certeza que la evaluación es similar independientemente del observador.

Del mismo modo, la división de categorías (disciplina alta, moderada y baja) no parece estar sustentada por criterios objetivos, incluso cuando existen métodos que son usados frecuentemente para ello (i. e., percentiles), así como procedimientos para garantizar hasta cierto punto la confiabilidad de la clasificación hecha luego de determinar el punto de corte (Dominguez-Lara, 2016b).

En este panorama, los resultados presentados por Sánchez Sanssoni (2015) son cuestionables no por el procedimiento estadístico realizado para el contraste de hipótesis, sino porque el instrumento empleado, al parecer, no permitiría obtener puntuaciones confiables ni inferencias válidas que permitan sostener los argumentos presentados. No obstante, cabe mencionar que cualquier estudio empírico de intervención debe complementarse con medidas de magnitud del efecto (ver Fritz, Morris, & Richler, 2012).

Con la presente reflexión no se pretende desacreditar el referido trabajo, dado que representa un esfuerzo que debe ser valorado, pero es necesario tomar conciencia de que deben ser evaluadas las propiedades psicométricas de todo instrumento utilizado, ya que de estas dependen en gran medida la validez de las conclusiones derivadas de los datos. Esto posee más trascendencia cuando se desea brindar evidencias científicas a favor de enfoques emergentes, ya que esos primeros resultados podrían (o no) alentar a la comunidad científica a continuar estudios sobre esos tópicos.

Referencias

Ato, M., López, J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología, 29*(3), 1038-1059.

Calderón-De la Cruz, G., Dominguez-Lara, S., & Navarro-Loli, J. (2016). ¿Es necesario describir los instrumentos empleados en investigaciones empíricas? *Anales de la Facultad de Medicina, 77*(2), 175-177.

Dominguez-Lara, S. (2016a). Validez de contenido usando la V de Aiken con intervalos de confianza: aportes a Rodriguez et al. *Archivos Argentinos de Pediatría, 114*(3), e221-e223.

Dominguez-Lara, S. (2016b). Confiabilidad de los puntos de corte en estudios empíricos. *Revista Cubana de Pediatría, 88*(4). Recuperado de http://www.revpediatria.sld.cu/index.php/ped/article/view/88/75

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect Size Estimates: Current Use, Calculations, and Interpretation. *Journal of Experimental Psychology: General, 141*(1), 2-18.

Merino-Soto, C., Calderón-De la Cruz, G., & Manzanares, E. (2016). Estudio comparativo del acuerdo y consistencia intercalificadores en el test gestáltico visomotor de Bender (2.ª ed.). *Revista Latinoamericana de Psicología, 48*(3), 175-182.

Sánchez Sanssoni, M. (2015). Efectos de un taller basado en la pedagogía del amor y la ternura para mejorar la disciplina en estudiantes de primer grado de primaria de una escuela arequipeña. *Educationis Momentum, 1*(1), 61-81.

Sireci, S., & Faulkner-Bond, M. (2014). Validity Evidence based on Test Content. *Psicothema, 26*(1), 100-107.