



CARTA AL EDITOR

Algunas recomendaciones sobre la elaboración de ítems

Effect size in regression analysis

Sergio Alexis Dominguez-Lara ^{1*}

¹ Universidad de San Martín de Porres, Perú

* Dirección de correspondencia: Instituto de Investigación de Psicología, Universidad de San Martín de Porres, Av. Tomás Marsano 242 (5to piso), Lima 34 – Perú; Teléfono: 0051988053909. Correo: sdominguezmpcs@gmail.com, sdominguezl@usmp.pe

Recibido: 21 de abril de 2017; Revisado: 29 de mayo de 2017; Aceptado: 02 de junio de 2017; Publicado Online: 03 de junio de 2017.

CITARLO COMO:

Dominguez-Lara, S. (2017). Algunas recomendaciones sobre la elaboración de ítems. *Interacciones*, 3(2), 95-XX. doi: 10.24016/2017.v3n2.61

Sr. Editor

En la actualidad, existe un interés creciente en la creación de instrumentos acordes a un contexto en particular, ya sea para fines de investigación o de aplicación profesional, pero en todo caso es necesario que cuenten con características psicométricas suficientes, confiabilidad y validez, para garantizar un uso responsable.

La validez hace referencia al grado en el cual las evidencias teóricas y empíricas permiten interpretar las puntuaciones en función del constructo de interés (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). En ese sentido, es condición necesaria que el proceso de construcción del instrumento garantice que las inferencias derivadas de las puntuaciones estarán fundamentadas, por lo que la elaboración de ítems es un punto crítico.

Entonces, cuando se desea construir un instrumento¹, el

investigador se preocupa por definir conceptualmente el *constructo* a evaluar, indicando a su vez si es unidimensional o multidimensional. Luego de ello, el proceso estándar de la elaboración de ítems tiene como objetivo lograr que el contenido de los ítems represente apropiadamente el constructo. Inclusive, la literatura ligada a la psicometría y construcción de instrumentos (p.e., Furr, 2011; Martínez, 1995) brinda pautas iniciales para el proceso de construcción de ítems, donde se destaca el formato de respuesta (tipo de formato, número de opciones, nombre de las etiquetas, etc.) y la redacción de ítems (contenido relevante, número de ítems, claridad, etc.).

No obstante, en ocasiones se observan problemas en esta etapa. Así, en escalas multidimensionales el contenido de algunos ítems no refleja de forma precisa el constructo general; o cada ítem contiene más de una situación (ítems con redacción compleja) que podrían confundir al evaluado (Hartley, 2014). Asimismo, en ocasiones los ítems al interior

¹ De aquí en adelante, se hará referencia a pruebas de respuesta típica (personalidad, actitudes, etc.), no de rendimiento.



de una escala tienen una estructura distinta (p.e., afirmaciones, preguntas, etc.), y no existe correspondencia lógica entre el contenido del ítem y las alternativas de respuesta que se brindan. Todas estas situaciones repercuten seriamente en la interpretación posterior de los resultados obtenidos, y más aún si se toman decisiones tomadas en base a estos.

Para ejemplificar los puntos mencionados anteriormente con relación a la elaboración de ítems, fue considerado el aporte de Gómez et al. (2015) sobre la construcción de una escala de actitudes hacia la obesidad.

Para comenzar, la *actitud* es definida en el manuscrito como “...creencias y sentimientos (positivos y negativos) sobre las personas, objetos y eventos...” (p. 277). Así, la *actitud hacia la obesidad* (AO) se definiría como creencias y sentimientos hacia la obesidad. En ese sentido, se espera que el contenido de cada uno de los ítems apunte a valorar las AO tal y como fue definida anteriormente. Pese a ello, algunos ítems de la dimensión *Preferencia/rechazo* (P/R), uno de los componentes de la AO, evalúan otros aspectos como conductas sociales (ítem 4) o apatía (ítem 6), que si bien podrían asociarse a la AO, no tienen que ver directamente con ella. Por ejemplo, si el primer ítem (*Con frecuencia, prefiero quedarme en casa, que salir a jugar al parque*) se lee aisladamente no refleja vínculo conceptual con la AO (Figura 1). Del mismo modo, la dimensión *preocupación/desinterés* (P/D) conceptualmente se asemeja más a la *preocupación por engordar* (PE), asociado a trastornos de la conducta alimentaria (p.e.,

anorexia nervosa), y si bien un puntaje alto en PE implicaría algún grado de inquietud de la persona por su peso, no necesariamente indica que su AO sea desfavorable.

Con respecto a la *complejidad del enunciado* en el proceso de redacción de ítems, esto puede suponer un problema ya que, al existir más de una idea, el respondiente solo puede reaccionar (esto es, escoger una opción de respuesta) a una de ellas. Como muestra, el ítem 1 indica *Con frecuencia, prefiero quedarme en casa, que salir a jugar al parque*. Cuando la persona responde una de las tres etiquetas disponibles (*totalmente de acuerdo, ni de acuerdo ni en desacuerdo, y totalmente en desacuerdo*) se podría determinar la intención de la respuesta. Si responde *totalmente en desacuerdo*, podría ser debido a: a) que no prefiera quedarse en su casa, y desee salir a jugar al parque; o b) que prefiera quedarse en su casa, pero no salir necesariamente al parque a divertirse. Evidentemente, la intención es que la respuesta brindada se corresponda solo la primera opción (a), pero la segunda también es posible.

Por otro lado, los ítems de las escalas de actitudes se componen habitualmente por afirmaciones, ante las cuales el evaluado expresará su grado de acuerdo o de desacuerdo. Aunque el formato ordinal de tres opciones es adecuado para la mayoría de los ítems presentados, algunos de estos son preguntas, por lo que responder de acuerdo con las etiquetas brindadas (p.e., *totalmente de acuerdo*) resultaría confuso para el evaluado, afectando la calidad de sus respuestas y, finalmente, la validez de la aplicación.

| Variable | Indicadores | Sub-indicadores |
|------------------------------|-------------------------|---|
| Actitud sobrepeso y obesidad | Preferencia/rechazo | 1. Con frecuencia prefiero quedarme en casa, que salir a jugar al parque. 2. A menudo no me llama la atención mucho hacer ejercicios físicos. 3. Me agrada salir a reuniones sociales con mis amigos. 4. Me siento más a gusto comiendo en la calle, que en mi casa. 5. Cuando tengo tiempo libre, prefiero mirar televisión en casa. 6. Cuando estoy en casa no me gusta ayudar en los quehaceres del hogar (limpiar, barrer, planchar, cortar pasto, etc). 7. Me agrada más la idea de comer comida rápida que las frutas 8. Cuando estoy en el colegio me gusta sentarme para conversar con mis amigos. |
| | Preocupación/desinterés | 9. Te sentirías preocupado en el caso de que subas de peso repentinamente? 10. Alguna vez me he sentido intranquilo por mi barriga? 11. Me preocupa mucho el hecho de no poder mantener mi peso ideal. 12. Actualmente consideras que estás subiendo de peso? 13. A menudo trato de no ingerir alimentos que me engorden. 14. Me preocupo mucho cuando ingiero alimentos ricos en grasas. 15. Últimamente siento que no estoy haciendo actividad física. |
| | Voluntad/Apatía | 16. Considero que tengo mucha voluntad para hacer ejercicios físicos. 17. En general, considero que tengo autocontrol cuando ingiero mis alimentos. 18. Pienso que soy capaz de mantener un estilo de vida saludable. 19. Creo que puedo subir de peso si me dejo vencer por la flojera. 20. Considero que los ejercicios físicos son importantes para mantener mi peso corporal. |

Figura 1. Estructura de la escala de actitud hacia la obesidad. Extraído de: Gómez et al. (2015).

Otro aspecto importante, pero en ocasiones no considerado, es la coherencia implícita entre el *contenido del ítem* y las *opciones de respuesta*. En los instrumentos que evalúan constructos como *personalidad* (y aspectos asociados), la inclusión de etiquetas de respuesta vinculadas con la *frecuencia* de la conducta (p.e., *Nunca, casi nunca, a menudo, algunas veces*, entre otros) es de utilidad porque si la persona reporta más conductas ligadas al constructo (expresadas en las opciones de mayor frecuencia, como *siempre* o *casi siempre*), podría inferirse que lo posee en mayor grado. Asimismo, en instrumentos donde el contenido de los ítems hacen referencia a reacciones emocionales de carácter situacional y se desea indagar sobre la *intensidad* de las mismas, las etiquetas de las respuesta poseen características que reflejan eso (p.e., *Nada, Algo, Bastante, Mucho*) (Dominguez-Lara, 2016), distinto de aquellas etiquetas que pretenden evaluar el grado de verdad que el evaluado considera para cada enunciado (p.e., *Totalmente verdadero*) (Merino-Soto & Salas-Blas, 2016) o el grado de acuerdo (p.e., *Muy de acuerdo*) con el contenido del ítem (Merino-Soto, Dominguez-Lara, & Fernández-Arata, 2017).

Concretamente, si las instrucciones del test hacen referencia a un momento particular (p.e., situación de examen), y el contenido del ítem es *Durante el examen me siento agitado*, resultaría incongruente que las opciones de respuesta sean: *Nunca, Casi nunca*, etc.

Sin embargo, la inclusión de esas palabras (p.e., *Nunca, casi nunca, nada, bastante*, entre otros) en el texto de los ítems podría afectar la respuesta del evaluado, sobre todo si existe cierta incompatibilidad entre el contenido del ítem y las etiquetas de respuesta.

Por ejemplo, el ítem 13 indica *A menudo trato de no ingerir alimentos que me engorden*. Entonces, si la persona responde *totalmente en desacuerdo* podría indicar que: a) la persona puede ingerir alimentos que la engorden o b) la persona trata de no ingerir alimentos que la engorden en menor frecuencia que *a menudo* (quizás una etiqueta de *pocas veces* en un escalamiento Likert más amplio). Un panorama más confuso resultaría si la persona responde *ni de acuerdo ni en desacuerdo* a ese ítem, ya que no se sabe si la posición neutra se enfoca en la frecuencia determinada por el ítem (*A menudo*), o en el contenido sustancial del reactivo (*tratar de no ingerir alimentos que engorden*).

Si bien es necesario para el psicólogo contar con instrumentos de medición que soporten su práctica profesional y que sean construidos en su contexto cultural, es más importante que estos cuenten con evidencias satisfactorias de validez, partiendo de una elaboración cuidadosa de sus ítems que permita valorar las puntuaciones apropiadamente.

Evidentemente, otros aspectos que no fueron profundizados en este manuscrito deben ser tomados en cuenta en la fase de análisis de las propiedades psicométricas para lograr los objetivos propuestos, como por ejemplo llevar a cabo análisis estadísticos según las características del ítem (p.e., *ordinal* o *continuo*) (Norman, 2010), implementar procedimientos congruentes con la evidencia de validez que se pretende hallar (p.e., análisis factorial para el estudio de la estructura interna) (Brown, 2006; Lloret-Segura, Ferreres-Traver, Hernández-Baeza, & Tomás-Marco, 2014), analizar si las dimensiones específicas configuran un constructo de orden superior (p.e., modelos jerárquicos) que justifique el uso de una puntuación total (Dominguez-Lara, & Rodriguez, 2017; Rodriguez, Reise, & Haviland, 2016) e, inclusive, evaluar el potencial sesgo asociado a los ítems invertidos (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Rozin, Berman, & Royzman, 2010) o a la deseabilidad social (Stöber, 2001).

Para concluir, los comentarios vertidos en la presente carta tienen como finalidad alertar a los investigadores y usuarios de tests psicológicos sobre algunas cuestiones prácticas que pueden ser omitidas en una etapa de elaboración de ítems, que como pudo apreciarse, es crítica en el proceso de construcción de un instrumento. Cabe precisar que algunos de los argumentos propuestos pueden no ser aplicables a toda la gama de instrumentos y constructos existentes en psicología, por lo cual deben tomarse con precaución.

CONFLICTO DE INTERESES

El autor expresa que no hay conflictos de interés al redactar el manuscrito.

REFERENCIAS

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington D.C.: American Psychological Association.
- Brown, T. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.
- Dominguez-Lara, S. (2016). Inventario de la Ansiedad ante Exámenes-Estado: análisis preliminar de validez y confiabilidad en estudiantes de psicología. *Liberabit*, 22(2), 219–228.
- Dominguez-Lara, S., & Rodriguez, A. (2017). Índices estadísticos de modelos bifactor. *Interacciones*. Publicación anticipada en línea. doi: 10.24016/2017.v3n2.51
- Furr, R.M. (2011). *Scale construction and psychometrics for social and personality psychology*. London: SAGE.
- Gómez, R., de Arruda, M., Sulla, J., Alvear, F., Urrea, C., González, C., Luarte, C., & Cossio, M. (2015). Construcción de un instrumento para valorar la actitud a la obesidad en adolescentes. *Revista Chilena de Nutrición*, 42(3), 277–283. doi: 10.4067/S0717-75182015000300008
- Hartley, J. (2014). Some thoughts on Likert-type scales. *International Journal of Clinical and Health Psychology*, 14(1), 83–86. doi: 10.1016/S1697-2600(14)70040-7

- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de psicología*, 30(3), 1151-1169. doi: 10.6018/analesps.30.3.199361
- Martínez, R. (1995). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- Merino-Soto, C., & Salas-Blas, E. (2016). Estructura de las motivaciones y dificultades percibidas para la investigación entre los docentes universitarios: Estudio preliminar. *Revista Interamericana de Psicología*, 50(2), 161 – 169.
- Merino-Soto, C., Dominguez-Lara, S., & Fernández-Arata, M. (2017). Validación inicial de una Escala Breve de Satisfacción con los estudios en estudiantes universitarios de Lima. *Educación Médica*, 18(1), 74 – 77. doi: 10.1016/j.edumed.2016.06.016
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5), 625-632. doi: 10.1007/s10459-010-9222-y
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879 – 903. doi: 10.1037/0021-9010.88.5.879
- Rodriguez, A., Reise, S.P., & Haviland, M.G. (2016). Evaluating bifactor models: calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137 – 150. doi: 10.1037/met0000045
- Rozin, P., Berman, L., & Royzman, E. (2010). Biases in the use of positive and negative words across twenty natural languages. *Cognition and Emotion*, 24, 536-548. doi: 10.1080/02699930902793462
- Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, 17(3), 222 – 232. doi: 10.1027//1015-5759.17.3.222