

Administración online versus lápiz-y-papel: funcionamiento diferencial de ítems de neuroticismo

*Online versus pencil-and-paper administration: differential items of neuroticism
functioning*

Facundo Juan Pablo Abal, Juan Franco Sánchez González, Gabriela Susana Lozzia,
Horacio Félix Attorresi
fabal@psi.uba.ar

Universidad de Buenos Aires (UBACyT2018 20020170100200BA y 20020170200001BA) y
Agencia Nacional de Promoción Científica y Tecnológica (PICT-2017-3226)

Eje temático: Evaluación psicológica y Técnicas y fundamentos de la Exploración
Psicológica

Resumen

La tendencia creciente hacia la informatización de las prácticas de evaluación psicológica ha propiciado que los investigadores y usuarios de tests recurran cada vez con más frecuencia a la recolección de datos *online* (Lozzia y otros, 2009). Esto permite acceder a muestras más numerosas, reducir los tiempos de administración y puntuación de las pruebas y ganar en comodidad para los evaluados. Pero también se esgrimen problemas de peso para desalentar su uso: el sesgo en la selección de los sujetos por su accesibilidad a Internet y el escaso control sobre el encuadre de la evaluación (Campos, Zucoloto, Bonafé, Jordani y Maroco, 2011; Weigold, Weigold, y Russell, 2013).

Recientemente Facundo J.P. Abal, Sofía E. Auné y Horacio F. Attorresi (2019) utilizaron la Teoría de Respuesta al ítem en la construcción de un Banco de ítems para la medición del Neuroticismo. Este rasgo describe la propensión a experimentar de manera intensa y/o recurrente una diversidad de afectos negativos (McCrae y Costa, 2010). La vinculación de este constructo con múltiples trastornos mentales y

físicos ha llevado a recomendar su medición en población general con fines preventivos (Tackett y Lahey, 2017; Widiger y Oltmanns, 2017). En este sentido, una administración *online* del Banco de Neuroticismo podría acarrear grandes ventajas para evaluar a más personas invirtiendo menos tiempo. No obstante, resulta menester garantizar que las respuestas proporcionadas a través de Internet son equivalentes a las recogidas con la versión de lápiz-y-papel.

El objetivo de este trabajo fue analizar el Funcionamiento Diferencial de los ítems (DIF - *Differential Item Functioning*) que componen el Banco de Neuroticismo en función del formato de administración usado para recolectar información: protocolo *online* versus protocolo con lápiz-y-papel. Se considera que un ítem presenta DIF si dos personas que respondieron distintos formatos muestran diferentes probabilidades de optar por alguna de las categorías de un ítem aun cuando tienen el mismo nivel de Neuroticismo (Camilli y Shepard, 1994).

Método

Participaron 742 personas de población general residentes en el Área Metropolitana de Buenos Aires. El 52.1% eran mujeres y la edad promedio fue de 29.5 ($DE = 11.32$; $Mín = 18$, $Máx = 82$). La mayoría de los participantes (63.3%) tenía como máximo un nivel secundario completo. Se utilizó un muestreo no-probabilístico por accesibilidad. Los participantes respondieron un protocolo que incluyó un cuestionario sociodemográfico y 24 elementos del Banco de Neuroticismo. El análisis de los datos se realizó con el programa IRTPRO 4.2 (Cai, Thissen y du Toit, 2011).

Resultados

Se realizaron análisis preliminares con la muestra completa: a) se verificaron los supuestos de independencia local y unidimensionalidad; y b) se constató el ajuste de los ítems al Modelo de Respuesta Graduada de F. Samejima (2016). En lo

atinente específicamente al estudio del DIF, se diferenció al 54.6% ($n = 405$) de los participantes que respondió el protocolo en versión lápiz-y-papel (grupo LP) y el 45.6% ($n = 337$) que contestó versión online (grupo OL) a través de *Google Forms*. Se corroboraron similares características sociodemográficas en ambos grupos. El análisis del DIF se llevó adelante siguiendo el procedimiento multietápico recomendado por L. Tay, A.W. Meade y M. Cao (2014). En la primera etapa se realizó una estimación simultánea de los parámetros de los ítems para los grupos LP y OL según el Modelo de Samejima. Esto permitió identificar 4 ítems con potencial DIF utilizando el test de Wald para detectar diferencias significativas entre los parámetros estimados para cada grupo ($p < .05$). Durante la segunda etapa se usaron como anclajes los 20 ítems que resultaron libres de DIF en el paso anterior. Esto permitió escalar a todos los ítems en una métrica común para poder examinar exclusivamente el presunto DIF de los 4 elementos señalados en la primera etapa. Como resultado final de la segunda etapa solo uno de los ítems presentó DIF significativo (χ^2 de Wald = 12.9; $gl = 4$, $p = .012$). La observación de los parámetros del ítem (“*Estoy tenso la mayor parte del tiempo*”) dejó en evidencia que el indicador mostró mayor capacidad discriminativa en el grupo LP.

Discusión y conclusión

Si bien uno de los ítems resultó con presencia de DIF no parece haber una razón que explique este resultado en términos de sesgo. Por ende, es posible sospechar que se trata de una falsa detección; esto es, un resultado debido al azar que lleva a cometer error de tipo I. No obstante, el ítem será excluido preventivamente y reevaluado en futuros estudios. Los ítems libres de DIF pasarán a formar parte del banco que servirá para el desarrollo futuro de un test adaptativo informatizado de Neuroticismo.

Palabras clave: neuroticismo, funcionamiento diferencial del ítems, test informatizado.

Abstract

The growing trend towards computerization of psychological evaluation practices has led researchers and test users to resort more and more frequently to online data collection (Lozzia et al., 2009). This allows access to more numerous samples, reduce administration times and test scores and gain in comfort for those evaluated. But weight problems are also mentioned to discourage their use: the bias in the selection of subjects for their accessibility to the Internet and the poor control over the evaluation setting (Campos, Zucoloto, Bonafé, Jordani & Maroco, 2011; Weigold, Weigold, & Russell, 2013).

Recently Facundo J.P. Abal, Sofía E. Auné y Horacio F. Attorresi (2019) used the Item Response Theory in the construction of an Item Bank for the measurement of Neuroticism. This trait describes the propensity to experience in an intense and/or recurrent manner a diversity of negative affects (McCrae & Costa, 2010). The linking of this construct with multiple mental and physical disorders has led to its measurement being recommended in the general population for preventive purposes (Tackett y Lahey, 2017; Widiger y Oltmanns, 2017). In this sense, an online administration of the Item Bank of Neuroticism could have great advantages to evaluate more people spending less time. However, it is necessary to ensure that the answers provided through the Internet are equivalent to those collected with the pencil-and-paper version.

The aim of this work was to analyze Differential Item Functioning (DIF) of the Item Bank of Neuroticism based on the administration format used to collect information: online protocol versus pencil-and-paper protocol. An item is considered to have DIF if two people who responded to different formats show different probabilities of choosing one of the categories of an item even when they have the same level of Neuroticism (Camilli & Shepard, 1994).

Method

742 people of general population resident in the Metropolitan Area of Buenos Aires participated. 52.1% were women and a mean age of 29.5 ($SD = 11.32$; $Min = 18$, $Max = 82$). The majority of participants (63.3%) had at most a complete secondary level. Non probability sampling methods (convenience sampling) was used. Participants responded to a protocol that included a socio-demographic questionnaire and 24 elements of the Item Bank of Neuroticism. Data analysis was performed with the IRTPRO 4.2 program (Cai, Thissen, & du Toit, 2011).

Results

Preliminary analyzes were carried out with the complete sample: a) the assumptions of local independence and one-dimensionality were verified; and b) the fit of the items to the Graded Response Model of F. Samejima (2016) was verified. With regard specifically to the study of DIF, 54.6% ($n = 405$) of the participants who responded to the protocol in pencil-and-paper version (PP group) and 45.6% ($n = 337$) who answered online version (OL group) through Google Forms were differentiated. Similar socio-demographic characteristics in both groups were corroborated. The DIF analysis was carried out following the multi-stage procedure recommended by L. Tay, A.W. Meade and M. Cao (2014). In the first stage, a simultaneous estimation of the parameters of the items for the PP and OL groups was carried out according to the Samejima's Model. This allowed us to identify 4 items with potential DIF using the Wald test to detect significant differences between the parameters estimated for each group ($p < .05$). During the second stage, the 20 items that were free of DIF in the previous step were used as anchors. This allowed scaling all the items in a common metric to be able to examine exclusively the presumed DIF of the 4 elements indicated in the first stage. As a final result of the second stage, only one of the items presented significant DIF (Wald $\chi^2 = 12.9$; $df =$

4, $p = .012$). The observation of the parameters of the item (*I am tense most of the time*) showed that the item was greater discriminative capacity in the PP group.

Discussion and conclusion

Although one of the items resulted in the presence of DIF, there seems to be no reason to explain this result in terms of bias. Therefore, it is possible to suspect that it is a false detection; that is, a result due to chance that leads to a type I error. However, the item will be excluded preventively and reassessed in future studies. DIF-free items will become part of the bank that will serve for the future development of a computerized adaptive Neuroticism test.

Keyword: neuroticism, differential item functioning, computerized test.

Referencias bibliográficas

- Abal, F.J.P., Auné, S.E. y Attorresi, H.F. (2019). Construcción de un banco de ítems de facetas de neuroticismo para el desarrollo de un test adaptativo. *Psicodebate*, 19 (1), pp. 31-50. doi: 10.18682/pd.v1i1.854.
- Cai, L., Thissen, D. y du Toit, S. (2011). *IRTPRO users guide*. Lincolnwood, IL: Scientific Software International.
- Camilli, G. y Shepard, L. (1994). *Methods for identifying biased test items*. Thousand Oaks: Sage.
- Campos, J.A.D.B., Zucoloto, M.L., Bonafé, F.S.S., Jordani, P.C. y Maroco, J. (2011). Reliability and validity of self-reported burnout in college students: A cross randomized comparison of paper-and-pencil vs. online administration. *Computers in Human Behavior*, 27(5), pp. 1875–1883. doi:10.1016/j.chb.2011.04.011.

- Lozzia, G., Abal, F.J.P., Blum, G.D., Aguerri, M., Galibert, M. y Attorresi, H. (2009). Tests Informatizados. Nuevos desafíos prácticos y éticos para la Evaluación Psicológica. *Summa Psicológica UST*, 6(1), pp. 135–148. doi: 10.18774/448x.2009.6.57.
- McCrae, R. y Costa P. (2010). *NEO Inventories professional manual*. Odessa, FL: Psychological Assessment Resources.
- Samejima, F. (2016). Graded Response Model. En W.J. van der Linden (Ed.). *Handbook of Item Response Theory, 1: Models* (pp. 95-108). Boca Raton: Chapman y Hall/CRC.
- Tackett, J.L. y Lahey, B.B. (2017). *The Oxford handbook of the five factor model*. New York: Oxford University Press.
- Tay, L., Meade, A.W. y Cao, M. (2014). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods*, 18 (1), pp. 3-46. doi: 10.1177/1094428114553062.
- Weigold, A., Weigold, I.K. y Russell, E.J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological Methods*, 18 (1), pp. 53–70. doi:10.1037/a0031607.
- Widiger, T.A. y Oltmanns, J.R. (2017). Neuroticism is a fundamental domain of personality with enormous public health implications. *World Psychiatry*, 16 (2), pp. 144–145. doi: 10.1002/wps.20411.