

XIV Congreso Internacional de Investigación y Práctica Profesional en Psicología. XXIX Jornadas de Investigación. XVIII Encuentro de Investigadores en Psicología del MERCOSUR. IV Encuentro de Investigación de Terapia Ocupacional. IV Encuentro de Musicoterapia. Facultad de Psicología - Universidad de Buenos Aires, Buenos Aires, 2022.

Funcionamiento Diferencial de ítems en la faceta Ansiedad: administración online versus lápiz-y-papel.

Steeb, Fernando y Lozzia, Gabriela.

Cita:

Steeb, Fernando y Lozzia, Gabriela (2022). *Funcionamiento Diferencial de ítems en la faceta Ansiedad: administración online versus lápiz-y-papel. XIV Congreso Internacional de Investigación y Práctica Profesional en Psicología. XXIX Jornadas de Investigación. XVIII Encuentro de Investigadores en Psicología del MERCOSUR. IV Encuentro de Investigación de Terapia Ocupacional. IV Encuentro de Musicoterapia. Facultad de Psicología - Universidad de Buenos Aires, Buenos Aires.*

Dirección estable: <https://www.aacademica.org/000-084/170>

ARK: <https://n2t.net/ark:/13683/eoq6/5he>

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

FUNCIONAMIENTO DIFERENCIAL DE ÍTEMS EN LA FACETA ANSIEDAD: ADMINISTRACIÓN ONLINE VERSUS LÁPIZ-Y-PAPEL

Steeb, Fernando; Lozzia, Gabriela

Universidad de Buenos Aires. Facultad de Psicología. Buenos Aires, Argentina.

RESUMEN

En la construcción de una escala para medir la faceta Ansiedad según el Modelo de los Cinco Factores, se analizó el Funcionamiento Diferencial de los ítems (DIF - Differential Item Functioning) según del formato de administración: online versus lápiz-y-papel. Se administraron 13 ítems tipo Likert a 742 personas de población general residentes en el área metropolitana de Buenos Aires (52.1% mujeres, edad media = 29.5 años y DE = 11.32 años). Un 54.6% (n = 405) de participantes respondieron el protocolo en lápiz-y-papel (grupo LP) y un 45.6% (n = 337) contestaron online (grupo OL) por Google Forms. Ambos grupos corroboraron similares características sociodemográficas. Se verificó la unidimensionalidad y el ajuste del Modelo de Respuesta Graduada de Samejima (2016) para todos los ítems. El análisis del DIF siguió el procedimiento multietápico de Tay et al. (2014). Únicamente el ítem Estoy tenso la mayor parte del tiempo presentó DIF significativo (χ^2 de Wald=11.3; $gl=3$, $p=.0102$) mostrando una localización más baja y mayor capacidad discriminativa para el grupo OL. Como no parece haber una explicación en términos de sesgo, podría tratarse de una falsa detección (error de tipo I). El ítem será excluido preventivamente y reevaluado en futuros estudios.

Palabras clave

Ansiedad - DIF - Test informatizado - Modelo de los Cinco Factores

ABSTRACT

DIFFERENTIAL ITEM FUNCTIONING OF ANXIETY FACET: ONLINE VERSUS PENCIL-AND-PAPER ADMINISTRATION

In the construction of a scale to measure the Anxiety facet according to the Five Factor Model, the Differential Item Functioning (DIF) was analyzed according to the administration format: online versus pencil-and-paper. Thirteen Likert-type items were administered to 742 people from the general population living in the metropolitan area of Buenos Aires (52.1% women, mean age = 29.5 years and SD = 11.32 years). 54.6% (n = 405) of participants answered the protocol in pencil-and-paper (LP group) and 45.6% (n = 337) answered online (OL group) by Google Forms. Both groups corroborated similar sociodemographic characteristics. The unidimensionality and fit of Samejima's (2016) Graded Response Model were verified for all items. DIF analysis

followed the multistage procedure of Tay et al. (2014). Only the item I am tense most of the time presented a significant DIF (Wald's $\chi^2=11.3$; $df=3$, $p=.0102$) showing a lower location and greater discriminative capacity for the OL group. As there does not seem to be an explanation in terms of bias, it could be a false detection (type I error). The item will be preemptively excluded and reevaluated in future studies.

Keywords

Anxiety - Differential Item Functioning - Computerized test - Five Factor Model

El Modelo de los Cinco Factores (FFM) se ha consolidado como uno de los abordajes factoriales de la personalidad más difundidos (Digman, 1990; Goldberg, 1993; McCrae y Costa, 2010) y de mayor apoyo empírico (Rammstedt et al., 2018; Kajonius y Johnson, 2019) en el campo de la evaluación psicológica. La propuesta central de este modelo reside en describir a la personalidad como un constructo que se puede organizar jerárquicamente en cinco rasgos generales. Las diversas combinaciones posibles de cada uno de estos rasgos es lo que daría lugar a las diferencias individuales manifiestas de cada persona. Estos factores generalmente se dividen en dominios denominados Neuroticismo, Extraversión, Apertura a la Experiencia, Amabilidad y Responsabilidad (McCrae y John, 1992). Los cinco dominios han dado muestras de su relevancia a la hora de intentar describir y explicar la personalidad (Piedmont, 1998).

El primero de estos factores en ser conceptualizado ha sido el Neuroticismo (Dumont, 2010). Si bien se han ensayado múltiples definiciones para este dominio, se observa en la literatura científica un notable consenso acerca de sus características (Jeronimus, 2015). En términos generales, puede definirse al Neuroticismo como la predisposición a sentir afectos negativos de manera intensa y/o prolongada (Goldberg, 1993; McCrae y Costa, 2010). Para el Neuroticismo, y también para los demás dominios, múltiples autores plantean un conjunto de facetas o subdimensiones a su interior que también son de relevancia a fines descriptivos de la personalidad, dado que permiten estudiar las diferencias individuales a un nivel de abstracción más definido y acotado en comparación con los dominios. Costa y McCrae (1995), los máximos exponentes del FFM, identifican seis facetas al interior

del Neuroticismo: Ansiedad, Hostilidad, Depresión, Autoconciencia, Impulsividad y Vulnerabilidad. Entre ellas, la faceta Ansiedad se destaca por ser una de las subdimensiones de mayor presencia y trascendencia en las diversas conceptualizaciones.

Costa y McCrae (1995) vinculan la Ansiedad con una predisposición a sentir temor de una forma inusualmente intensa o persistente. Las personas con altos niveles de Ansiedad tienden a pensar que algo peligroso está por ocurrir, y experimentan frecuentemente nervios, tensión y/o preocupación (Soto y John, 2016; Watson et al., 2019). Además, muestran una fuerte inclinación a intentar evitar fracasos a toda costa, de modo que no suelen involucrarse en actividades competitivas ni a tomar riesgos. En contraposición, quienes muestran bajos niveles en este rasgo son generalmente personas que no piensan que las cosas pueden salir mal y que experimentan con mayor frecuencia sensaciones de tranquilidad o serenidad (Taylor y DeBruin, 2013).

Actualmente, se está desarrollando en el ámbito local una escala para medir la faceta Ansiedad basada en la definición del FFM. Esta escala será parte de un banco de ítems para la medición del Neuroticismo (Abal et al., 2019). Su construcción implica llevar adelante estudios de confiabilidad y validez pero también determinar cuestiones vinculadas a la forma de administración del instrumento. Esto último tiene su fundamento en que en la actualidad es cada vez más frecuente el uso de plataformas virtuales online como vía de recolección de datos para evaluaciones psicológicas (Lozzia et al., 2009). Esta modalidad tiene ciertas ventajas, tales como la comodidad de la administración, el fácil acceso a muestras más numerosas y la disminución de los tiempos de administración y puntuación de las pruebas. No obstante, su uso también conlleva el riesgo de caer en sesgos debido a la selección de los sujetos por su accesibilidad a Internet y por tener un escaso control sobre el encuadre de la evaluación (Campos et al., 2011; Weigold et al., 2013).

Frente a esta situación, es imprescindible verificar que los datos recolectados a través de Internet sean equivalentes a los obtenidos con la versión de lápiz-y-papel. Por ello, en el marco de la construcción de una escala para medir la faceta Ansiedad, este trabajo busca determinar si las respuestas a este instrumento recogidas mediante una administración clásica de lápiz-y-papel pueden considerarse homólogas a las obtenidas a través de una administración online. Se trata de identificar si los ítems administrados de distinta forma presentan un Funcionamiento Diferencial (*Differential Item Functioning*, DIF). Un ítem presenta DIF si dos personas que tienen el mismo nivel de Ansiedad muestran diferentes probabilidades de optar por alguna de las categorías de un ítem sólo porque respondieron distintos formatos de administración (Camilli y Shepard, 1994). El análisis del DIF constituye un aporte actual de la psicometría de gran importancia en la etapa de construcción y validación de tests (Mallinckrodt et al., 2016; Zumbo, 2007).

Concretamente, el objetivo de este trabajo es analizar el Funcionamiento Diferencial de los ítems que componen una esca-

la para medir la faceta Ansiedad según el FFM en función del formato de administración usado para recolectar información: protocolo online versus protocolo con lápiz-y-papel.

Método

Participantes

Colaboraron 742 personas (52.1% género femenino) de población general residentes en el área metropolitana de Buenos Aires, con edades comprendidas entre 18 y 82 años. Se utilizó un muestreo no-probabilístico por accesibilidad tanto para la versión en lápiz y papel como para la versión online. En la tabla 1 se detallan las características sociodemográficas tanto para el grupo que respondió el protocolo en lápiz-y-papel (LP) como para el grupo que lo hizo online (OL). Se observó una notable similitud entre ambos grupos para género, edad, nivel socioeconómico y nivel educativo.

Tabla 1

Caracterización de la muestra

		LP (n = 405)	OL (n = 337)
Género	Femenino	52.2%	51.9%
	Varonil	47.8%	48.1%
Edad	Media (DE)	29.6 (11.4)	29.4 (11.1)
	Mínimo - Máximo	18 - 82	18 - 81
Nivel socioeconómico	Nivel medio	82.5%	83.1%
Nivel educativo	Secundario completo	63.4%	63.6%

Instrumentos

Cuestionario sociodemográfico. Diseñado *ad hoc* para los fines de esta investigación. Indaga variables tales como género, edad, estado civil, lugar de nacimiento, lugar de residencia, situación laboral y nivel de estudios alcanzado.

Ítems de la faceta Ansiedad. Son 13 ítems que operacionalizan la faceta Ansiedad en el marco del FFM. Presentan un formato de respuesta de escala Likert de cuatro opciones (En desacuerdo, Ligeramente en desacuerdo, Ligeramente de acuerdo y De acuerdo). Los ítems pretenden dar cuenta del grado en que una persona tiende a sentir temor, preocupación y/o tensión de manera inusualmente intensa y/o asidua. La consistencia interna de todos los reactivos indicó un alfa de Cronbach .87 en la muestra de este estudio.

Procedimiento

Los participantes que contestaron el protocolo en formato lápiz-y-papel lo hicieron individualmente bajo la supervisión de un profesional y sin límite de tiempo. En la administración online se utilizó la plataforma Google Forms para configurar una versión

informatizada del protocolo, el cual se difundió vía correos electrónicos y redes sociales. Se incluyó un consentimiento informado, información sobre los propósitos de la investigación y aclaraciones sobre las condiciones de anonimato y confidencialidad de la participación. Los evaluados recibieron también datos de contacto del director de la investigación para eventuales consultas. El diseño de este estudio fue evaluado y aprobado por el Comité de Conductas Responsables de la Facultad de Psicología de la Universidad de Buenos Aires.

Análisis de datos

El análisis de los datos se realizó con el programa IRTPRO 4.2 (Cai et al., 2011). Para verificar los supuestos de independencia local y unidimensionalidad, se calcularon índices de proximidad a la unidimensionalidad, se realizó un Análisis Factorial Exploratorio (AFE) y se estimó el estadístico χ^2 de Chen y Thissen (1997). Además, se constató el ajuste de los ítems al Modelo de Respuesta Graduada (MRG) de Samejima (2016). Para el estudio del DIF se siguió el procedimiento multietápico recomendado por Tay et al. (2014).

Resultados

Análisis preliminar de los ítems

Verificación de los supuestos del MRG. Se verificó el supuesto de unidimensionalidad del constructo a partir de los índices de proximidad a la unidimensionalidad ($ECV = .91$, $MIREAL = .15$) y de un AFE ($KMO = .93$; Prueba Bartlett, $\chi^2 = 2911.5$; $gI = 78$; $p < 0.001$) con la implementación óptima del análisis paralelo de Horn que indicó la extracción de un único factor dominante que describe un 63.6% de la varianza. También se corroboró la independencia local de los ítems ($\chi^2 < 10$).

Estimación y ajuste del MRG. Cumplidos los supuestos, se constató el ajuste de los ítems al Modelo de Respuesta Graduada (MRG) de Samejima (2016) usando el índice $S-\chi^2$ de Orlando y Thissen (2003). Todos los ítems se ajustaron al MRG ($p > .01$).

Análisis del DIF

Estudio del DIF: Primera etapa. Se segmentó la muestra en dos grupos: Lápiz/Papel y Online. Se realizó para cada grupo una estimación simultánea pero independiente de los parámetros de los ítems según el MRG. A partir del test de Wald, que examina diferencias significativas entre los parámetros estimados para cada grupo, se identificó un ítem con potencial DIF ($p < .05$) y 12 ítems libres de DIF.

Estudio del DIF: Segunda etapa. Los 12 ítems libres de DIF se usaron como anclaje para escalar a todos los ítems en una métrica común. Esto permitió examinar exclusivamente el presunto DIF del reactivo señalado en la primera etapa. Como resultado final de la segunda etapa, solo este ítem (*Estoy tenso la mayor parte del tiempo*) presentó DIF significativo (χ^2 de Wald = 11.3; $gI = 3$, $p = .0102$), mostrando una localización más baja y mayor

capacidad discriminativa para el grupo que respondió online. En la tabla 2 se detallan los resultados del análisis para el ítem detectado con DIF.

Tabla 2

Ítem detectado con DIF y sus parámetros estimados según MRG.

Ítem		Calibración			
		a	b1	b2	b3
<i>Estoy tenso la mayor parte del tiempo</i>	Grupo LP	1.71	0.08	1.06	2.22
	Grupo OL	1.82	-0.17	0.70	1.67

Nota. a = Parámetro de discriminación del Modelo Logístico de dos parámetros. b = Parámetro de localización del Modelo Logístico de dos parámetros.

Discusión

La evaluación psicológica a partir de test administrados de manera online es una práctica cada vez más habitual. Como ya fue mencionado, esta forma de administración cuenta con muchas ventajas, pero también puede implicar la presencia de sesgos que ameritan ser evaluados para garantizar la validez de los puntajes obtenidos y su comparabilidad con los resultados recogidos a través de la administración tradicional de lápiz y papel (Lozzia et al., 2009). En este sentido, se llevó adelante el estudio del DIF en los ítems de la faceta Ansiedad ya que es un procedimiento que brinda evidencia empírica de que las mediciones obtenidas son invariantes con independencia del formato de administración (Zumbo, 2007).

Solo uno de los 13 ítems mostró presencia de DIF (*Estoy tenso la mayor parte del tiempo*). Para las personas que respondieron de forma online, este ítem mostró mayor capacidad discriminativa y requirió menor nivel de ansiedad para acordar con cada una de las categorías de respuesta. Sin embargo, no hay una razón evidente que explique este resultado en términos de sesgo. Por lo tanto, podría tratarse de una falsa detección. Es decir, se observó un resultado poco probable que es debido al azar, y por ello, llevaría a cometer error de tipo I (rechazar la hipótesis nula cuando en realidad es verdadera). No obstante, el ítem fue excluido preventivamente para ser evaluado en próximos estudios con el fin de precisar si la diferencia observada en este trabajo es genuina o si se provoca artificialmente por error de muestreo. Futuros estudios podrían llevarse a cabo teniendo en cuenta otras características sociodemográficas de los participantes a la hora de estudiar el funcionamiento diferencial de los ítems utilizados. Por ejemplo, también es importante estudiar un posible DIF según el género o por grupo etario.

En definitiva, este trabajo ha constituido un aporte a las evidencias de validez de estos nuevos ítems desarrollados para evaluar la faceta Ansiedad. Los ítems que fueron considerados como libres de DIF se incluyeron dentro de un banco de ítems que servirá para el desarrollo futuro de un test adaptativo informatizado de Neuroticismo.

BIBLIOGRAFÍA

- Abal, F.J.P., Auné, S.E., y Attorresi, H. (2019) Construcción de un banco de ítems de facetas de neuroticismo para el desarrollo de un test adaptativo. *Psicodebate*, 19(1), 31-50. <https://doi.org/10.18682/pd.v19i1.854>
- Cai, L., Thissen, D., y du Toit, S. (2011) *IRTPRO Users Guide*. Scienti?c 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), 1875-1883. <https://doi.org/10.1016/j.chb.2011.04.011>
- Chen, W.H., y Thissen, D. (1997) Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265-289. <https://doi.org/10.3102/10769986022003265>
- Costa, P.T., y McCrae, R.R. (1995) Domains and Facets: Hierarchical Personality Assessment Using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, 64(1), 21-50. https://doi.org/10.1207/s15327752jpa6401_2
- Digman, J.M. (1990) Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440. <https://doi.org/10.1146/annurev.ps.41.020190.002221>
- Dumont, F. (2010) *A History of Personality Psychology: Theory, Science, and Research from Hellenism to the Twenty-first Century*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511676093>
- Goldberg, L.R. (1993) The structure of phenotypic personality traits. *American Psychologist*, 48(1), 26-34. <https://doi.org/10.1037/0003-066X.48.1.26>
- Mallinckrodt, B., Miles, J.R., y Recabarren, D.A. (2016) Using focus groups and Rasch item response theory to improve instrument development. *The Counseling Psychologist*, 44(2), 146-194. <https://doi.org/10.1177/0011000015596437>
- McCrae, R.R., y Costa, P.T. (2010) *NEO Inventories Professional Manual*. Psychological Assessment Resources.
- McCrae, R.R., y John, O.P. (1992) An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, 60(2), 175-215. <https://doi.org/10.1111/j.1467-6494.1992.tb00970.x>
- Jeronimus, B.F. (2015) *Neuroticism & Environments: A story about emotional (in)stability* [Tesis doctoral, Universidad de Groningen]. [S.l.]: [S.n.]. <https://bit.ly/3Axywe5>
- Kajonius, P.J., y Johnson, J.A. (2019) Assessing the Structure of the Five Factor Model of Personality (IPIP-NEO-120) in the Public Domain. *Europe's Journal of Psychology*, 15(2), 260-275. <https://doi.org/10.5964/ejop.v15i2.1671>
- 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), 135-148. <https://doi.org/10.18774/448x.2009.6.57>
- Orlando, M., y Thissen, D. (2003) Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, 27(4), 289-298. <https://doi.org/10.1177/0146621603027004004>
- Piedmont, R.L. (1998) *The Revised NEO Personality Inventory: Clinical and Research Applications*. Plenum Press. <https://doi.org/10.1007/978-1-4899-3588-5>
- Rammstedt, B., Lechner, C., y Danner, D. (2018) Relationships between Personality and Cognitive Ability: A Facet-Level Analysis. *Journal of Intelligence*, 6(2), 28. <https://doi.org/10.3390/jintelligence6020028>
- Samejima, F. (2016) Graded Response Model. En W. J. van der Linden (Ed.), *Handbook of Item Response Theory, Volume 1: Models* (pp. 95-108). Boca Raton: Chapman y Hall/CRC.
- Soto, C.J., y John, O.P. (2016) The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1), 117-143. <https://doi.org/10.1037/pspp0000096>
- Taylor, N., y de Bruin, G.P. (2013) The Basic Traits Inventory. En S. Laher & K. Cockcroft (Eds.), *Psychological Assessment in South Africa: Research and Applications* (pp. 232-243). Wits University Press. <https://doi.org/10.18772/22013015782.21>
- Tay, L., Meade, A.W., y Cao, M. (2014) An Overview and Practical Guide to IRT Measurement Equivalence Analysis. *Organizational Research Methods*, 18(1), 3-46. <https://doi.org/10.1177/1094428114553062>
- Watson, D., Nus, E., y Wu, K.D. (2019) Development and validation of the Faceted Inventory of the Five-Factor Model (FI-FFM). *Assessment*, 26(1), 17-44. <https://doi.org/10.1177/1073191117711022>
- 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), 53-70. <https://doi.org/10.1037/a0031607>
- Zumbo, B.D. (2007) Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223-233. <https://doi.org/10.1080/15434300701375832>