

XIII Congreso Internacional de Investigación y Práctica Profesional en Psicología. XXVIII Jornadas de Investigación. XVII Encuentro de Investigadores en Psicología del MERCOSUR. III Encuentro de Investigación de Terapia Ocupacional. III Encuentro de Musicoterapia. Facultad de Psicología - Universidad de Buenos Aires, Buenos Aires, 2021.

El uso de modelos de la teoría de respuesta al ítem en Test de Personalidad.

Sanchez Gonzalez, Juan Franco, Piccini, Stefania y Abal, Facundo Juan Pablo.

Cita:

Sanchez Gonzalez, Juan Franco, Piccini, Stefania y Abal, Facundo Juan Pablo (2021). *El uso de modelos de la teoría de respuesta al ítem en Test de Personalidad. XIII Congreso Internacional de Investigación y Práctica Profesional en Psicología. XXVIII Jornadas de Investigación. XVII Encuentro de Investigadores en Psicología del MERCOSUR. III Encuentro de Investigación de Terapia Ocupacional. III Encuentro de Musicoterapia. Facultad de Psicología - Universidad de Buenos Aires, Buenos Aires.*

Dirección estable: <https://www.aacademica.org/000-012/130>

ARK: <https://n2t.net/ark:/13683/even/wah>

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>.

EL USO DE MODELOS DE LA TEORÍA DE RESPUESTA AL ÍTEM EN TEST DE PERSONALIDAD

Sanchez Gonzalez, Juan Franco; Piccini, Stefania; Abal, Facundo Juan Pablo
ANPCyT - Universidad de Buenos Aires. Buenos Aires, Argentina.

RESUMEN

El objetivo de este trabajo fue plantear un ordenamiento histórico que permita explicar el desarrollo de la aplicación de la Teoría de Respuesta al Ítem (TRI) en tests de personalidad. Se distinguieron tres etapas históricas. Una Primera etapa se distingue por ser un periodo en el que se estudió el ajuste de los modelos a los datos de la personalidad limitándose al análisis de las propiedades psicométricas de los ítems. La Segunda etapa se caracteriza por una tendencia a capitalizar los beneficios de la TRI e identificar las diferentes complicaciones que se presentaban en la práctica. La Tercera etapa refleja el afianzamiento del conocimiento adquirido en etapas anteriores, que permitió nuevas estrategias para superar los obstáculos y el avance en la construcción de Bancos de ítems y Test Adaptativos Informatizados. Las características de cada etapa se asientan en las limitaciones teóricas y tecnológicas de cada periodo. Resulta dificultoso fechar estas etapas debido a la dispar difusión de la TRI internacionalmente. Mas allá de esto, puede afirmarse que la TRI plantea un fructífero panorama a futuro y en el campo de la medición de la personalidad esto es sumamente importante.

Palabras clave

TRI - Personalidad - Evaluación psicológica - Test de personalidad

ABSTRACT

THE USE OF ITEM RESPONSE THEORY MODELS IN PERSONALITY TESTS

The aim of this paper was to propose a historical order to explain the development of the application of Item Response Theory (IRT) in personality tests. Three historical stages were distinguished. The first stage is distinguished by a period in which the adjustment of the models to the personality data was studied, limiting itself to the analysis of the psychometric properties of the items. The second stage is characterized by a tendency to capitalize the benefits of IRT and to identify the different complications that arose in practice. The third stage reflects the consolidation of the knowledge acquired in previous stages, allowing new strategies to overcome obstacles and progress in the construction of Item Banks and Computerized Adaptive Tests. The characteristics of each stage are based on the theoretical and technological limitations of each period. It is difficult to date these stages due to the disparate diffusion of IRT internationally. Beyond this, it can be affirmed that IRT offers a fruitful outlook

for the future and in the field of personality measurement this is extremely important.

Keywords

IRT - Personality - Psychological - Assessment - Personality test

Introducción

La Teoría de Respuesta al Ítem (TRI) surge como un modelo de medición superador de las limitaciones que presentaba la Teoría Clásica de los Test (TCT). Su nacimiento se ubica a partir de la tesis doctoral de Lord (1952) "*A theory of test scores*", aunque pueden rastrearse publicaciones anteriores como antecedentes formales e incipientes de la TRI (Thurstone, 1927; Richardson, 1936; Lawley, 1943; Tucker, 1946). Aun así, debido a la complejidad de los cálculos no fue posible hasta la década de los 70, con el avance de la informática y el surgimiento de softwares estadísticos, que se pudo poner en práctica la aplicación de los modelos de la TRI (Abal et al., 2007).

La principal característica de la TRI es tomar como unidad de análisis a los ítems, y no la escala completa como ocurre en la TCT, vinculando el nivel de rasgo latente que un individuo posee con la probabilidad de dar la respuesta clave en el reactivo (Baker y Kim, 2017). Las ventajas que presenta la TRI por sobre la TCT redundan en que ha permitido elaborar instrumentos con propiedades invariantes entre poblaciones y mediciones invariantes más allá de la composición de la prueba (Muñiz, 2018). Estas características propician formas de administración de los instrumentos más flexibles y eficientes que las tradicionales, como las propiciadas por Tests Adaptativos Informatizados desarrollados a partir de Bancos de Ítems (Nieto et al., 2017).

La TRI significó una revolución a la hora de construir y administrar test utilizados en el campo de la evaluación psicológica de aptitudes cognitivas y las pruebas educativas (Muñiz, 2018). No obstante, su aplicación a pruebas que miden rasgos de personalidad tuvo que superar múltiples obstáculos. Por este motivo, se observa una notoria desproporción de artículos instrumentales en los que se aplica esta teoría a la construcción de test de rendimiento máximo comparado con producciones que aplican TRI para modelizar aspectos de la personalidad. En una revisión bibliográfica realizada por Abal et al (2010) se identifica que esta desproporción puede deberse a una conjunción de factores históricos que entorpecieron su difusión y factores operativos que dificultaban la utilización de la TRI por características inhe-

rentes a las variables de personalidad.

En este trabajo se propone un recorrido histórico que permita dimensionar los avances que ha tenido la creciente implementación de la TRI en su aplicación a escalas de personalidad. El desarrollo se ha organizado en tres etapas que permiten, en términos generales, sistematizar los objetivos instrumentales que perseguían los psicometristas en cada época para llevar a la práctica el uso de la TRI.

Desarrollo

Primera etapa: las aplicaciones exploratorias

Las primeras aplicaciones de la TRI a test de personalidad estuvieron condicionadas fuertemente por los modelos que habían sido diseñados previamente para la medición de variables relacionadas al ámbito educativo. Por esta razón, en los primeros momentos los psicometristas volcaron su atención en el uso de modelos para ítems dicotómicos, los cuales son los que con mayor frecuencia son empleados en test de ejecución máxima. Surgieron así aplicaciones de la TRI a pruebas como el Cuestionario de Personalidad de Eysenck (EPQ) o el Inventario Multifásico de Personalidad de Minnesota (MMPI-2) (e.g. Carter & Wilkinson, 1984), cuyos ítems presentaban un formato de respuesta dicotómico acorde con las características de los modelos logísticos iniciales.

Sin embargo, los ítems con dos opciones de respuesta no resultan tan frecuente en las medidas de la personalidad. En su mayoría, los test de personalidad se caracterizan por presentar un formato de respuesta politómica ordenada. Esto llevó a que los psicometristas tiendan a dicotomizar las categorías de las respuestas tipo Likert (considerando como *En desacuerdo* los valores inferiores de la escala politómica y *De acuerdo* los valores superiores). Este procedimiento era habitual cuando no se disponía aun de modelos politómicos (e.g. Bock, 1972). Posteriormente, se fueron desarrollando distintos modelos aplicables a ítems de respuesta politómica. El primer modelo fue el Modelo de Respuesta Graduada de Samejima (1969), que si bien fue propuesto inicialmente para test de ejecución máxima, sus características permitieron su aplicación en ítems con escalas tipo Likert (e.g. Bejar, 1977).

Esta etapa se caracterizó por una aplicación exploratoria de los modelos de la TRI a medidas de la personalidad (e.g. Ferrando, 1994; Reise y Waller, 1990). Los objetivos propios de este momento se acotaban a estudiar el ajuste de los modelos a los datos de la personalidad limitándose exclusivamente al análisis de las propiedades psicométricas de los ítems (Abal et al., 2010). Esto inauguró un lento pero constante incremento en las aplicaciones de la TRI que permitió comprender cómo debían interpretarse los parámetros de los modelos al analizar ítems de personalidad (Abal et al., 2009).

Segunda etapa: aparición de limitaciones y avances

Durante esta segunda etapa se incrementaron los trabajos que hacían uso de los modelos de la TRI para calibrar ítems pertenecientes a diferentes escalas de personalidad y así lo evidencian numerosas investigaciones (e.g. McBride, 2001; Spence et al., 2012). No obstante, esta segunda etapa se caracteriza por una tendencia a capitalizar los beneficios de la TRI en pos de alcanzar una mayor profundidad en aspectos de la medición que la TCT no permitía. Un ejemplo claro de esto es el análisis de la Función de Información del Test (FIT) considerando las características del constructo psicológico evaluado. Mientras las recomendaciones tradicionales sostenían la necesidad de alcanzar una FIT elevada y simétrica respecto de la media del rasgo, pronto se comprendió que este resultado era esperable sólo si el constructo describía personalidad normal. Por el contrario, los instrumentos que evalúan rasgos patológicos de la personalidad presentan elevaciones de la FIT en niveles extremos de la variable, en donde resulta más preciso para discriminar las diferencias individuales entre los individuos evaluados (Pérez Sutil et al., 2019).

Los avances de esta etapa se vinculan principalmente con la identificación de diferentes complicaciones que se presentaban en la práctica. En efecto, la aplicación de la TRI se sostiene en supuestos rigurosos que solían descuidarse cuando se analizaban escalas de manera exploratoria. El supuesto de unidimensionalidad del rasgo latente era necesario para la aplicación de gran parte de los modelos de la TRI (Morizot et al., 2007). Para este período, los modelos multidimensionales estaban desarrollados a nivel teórico pero su implementación práctica estaba limitada por el avance de la tecnología. El ajuste de modelos unidimensionales entraba en conflicto con el hecho de que las variables de personalidad son, en su mayoría, inherentemente multidimensionales. Para dar solución provisional a este problema en esta etapa se tendió a utilizar los modelos a dimensiones aisladas factorialmente, tratando por separado y como unidimensional a cada rasgo identificado en un constructo multidimensional (Frale y et al., 2000).

Otro de los supuestos que con mayor facilidad era violado al aplicar la TRI en escalas de personalidad es el de Independencia Local de los ítems (Abal, 2013). Es habitual que el uso de los modelos de la TRI sea limitado al aplicarse a las escalas que fueron diseñadas y validadas usando como marco metodológico la TCT. Esto se debe a que es frecuente la utilización de ítems redundantes que aumentan artificialmente la consistencia interna de la escala. No obstante, desde la perspectiva de la TRI no es posible indicar que estos ítems sean independientes estadísticamente entre sí (Abal et al., 2010).

Con el avance en el uso de los modelos politómicos de la TRI se inició el estudio de la contribución que tienen las escalas Likert en test de personalidad en la calidad psicométrica del ítem y del instrumento. Es decir, se empezaron a ensayar y comparar diferentes formatos de respuesta con el fin de concluir

cuál resultaba más adecuado para optimizar la medición (e.g. García Cueto et al., 2003; Kamp Denegri, 2006). Con base en la evidencia acumulada, se ha definido que cuatro opciones de respuesta pueden resultar una cantidad óptima para garantizar un equilibrio entre el grado de ajuste al modelo y la precisión de la medida (Abal, et al., 2018). No obstante, estos estudios continúan actualmente porque los hallazgos no son concluyentes. También en este período comenzaron a expandirse los estudios del Funcionamiento Diferencial de Ítems (DIF) aplicado a reactivos de escalas de personalidad. Este tipo de estudios consiste en evaluar si las diferencias encontradas entre grupos específicos (denominados de referencia y focal) se deben a diferencias genuinas en el rasgo o son generadas artificialmente por los ítems que componen el instrumento presentando un funcionamiento diferencial (Abal, 2013). Los primeros estudios contemplaban comparaciones entre grupos a partir de características sociodemográficas básicas tales como género o grupo étnico (Escorial y Navas, 2007; Sheppard et al., 2006). Sin embargo, posteriormente aparecieron aplicaciones en las que se examinaba la calidad de la adaptación de instrumentos de una cultura a otra (Johnson et al., 2008) o se comparaban los resultados entre la administración de un test en formato online *versus* su formato tradicional lápiz-papel (e.g. Meade et al., 2007; Abal et al., 2020).

Tercera etapa: una consolidación llevada al campo aplicado

Esta última etapa se caracteriza por una consolidación en el conocimiento adquirido en la aplicación de la TRI a las escalas de personalidad. En este período se empezaron a desarrollar algunas estrategias que permiten superar los obstáculos ocasionados por los supuestos rigurosos de la TRI. Por ejemplo, con el fin de evitar los problemas asociados a la violación del supuesto de independencia local en los últimos años algunos autores se volcaron a la construcción de instrumentos con ítems originales en lugar de utilizar los desarrollados en el marco de la TCT (e.g. Abal et al., 2019; Nieto et al., 2017). Asimismo, el progreso en el desarrollo de los software también propició el surgimiento de aplicaciones de modelos multidimensionales a las escalas que miden la personalidad (e.g. Colledani et al., 2019). Esto permite a los investigadores analizar las interrelaciones existentes en la estructura interna del constructo sin tener que recortarlo operativamente para su estudio psicométrico.

En esta línea, es interesante añadir que también se empezaron a aplicar modelos no paramétricos de la TRI a la medición de la personalidad, como el modelo de Mokken (1971) (e.g. Bech et al., 2016; Maydeu Olivares, 2005; Watson et al., 2007). Estos modelos son una alternativa basada solo en restricciones de orden, por lo que son más flexibles y menos restrictivos que los modelos paramétricos usualmente empleados en la TRI (Luzardo, 2017). Uno de los aspectos en los que se profundizó en este período es el impacto de la deseabilidad social o distorsión en las respuestas de los test de personalidad. Con la implementación de los modelos probabilísticos de la TRI, se ha podido estudiar a través

de diferentes métodos la detección de la distorsión en las respuestas de los evaluados. La identificación de patrones de respuestas atípicos se analizó al examinar el ajuste de la secuencia de opciones elegidas por un evaluado con las predicciones del modelo (Brown y Harvey, 2003; Stark et al., 2012). Por otro lado, Ferrando (2008) aplicó un modelo multidimensional de la TRI en el EPQ pudiendo analizar la relación de la escala de Sinceridad con la estimación del resto de las dimensiones del modelo.

La construcción de Bancos de Ítems que permiten evaluar diferentes rasgos de la personalidad es una de las ventajas más sobresalientes de la TRI y un campo de desarrollo muy prometedor. En esta línea de investigaciones se han generado diversos Bancos que operacionalizan distintos rasgos de personalidad (e.g. Abal et al., 2019; Stepp, 2012; Highhouse et al., 2016). Sobre la base de estos Bancos de ítems es posible la aplicación de los ítems que los componen como Test Adaptativo Informatizado. Esta estrategia permite el ahorro de ítems y tiempo de administración mejorando la eficiencia de las evaluaciones sin la pérdida de validez de contenido y manteniendo la confiabilidad de las puntuaciones. En los últimos años se han ido implementando cada vez con más frecuencia este tipo de estrategias que, sumado a la creciente integración de las tecnologías en la evaluación psicológica, plantean un panorama interesante de cara al futuro (e.g. Luo et al., 2020).

Discusión

Es innegable el crecimiento que ha tenido en los últimos años la aplicación de la TRI en la medición de rasgos de la personalidad. No obstante, está muy lejos de ser usada con la misma asiduidad con que se usa en test de rendimiento máximo. La presente revisión histórica busca retratar la evolución que han tenido y los obstáculos que se han superado desde aquellas primeras aplicaciones exploratorias.

Esta división en etapas se realizó a fines de facilitar la comprensión de los desarrollos históricos de la TRI en el área de la personalidad. Las características principales de cada etapa se asientan en las limitaciones teóricas y tecnológicas imperantes en cada período. Resulta dificultoso fechar estas etapas en virtud de que la difusión de la TRI ha sido despareja en los distintos países del mundo y esto condiciona los años de publicación de los artículos.

Teniendo en cuenta lo expuesto anteriormente, puede afirmarse que la TRI plantea un fructífero panorama a futuro, ya que son infinitas las posibilidades y nuevos modelos que esta teoría permitiría desarrollar. En el campo de la evaluación de la personalidad esto es sumamente importante, debido a que esto supone la mejora y refinamiento de los instrumentos que permiten medir los rasgos de personalidad. Esto redundará en la profundización del conocimiento sobre estos constructos y en un perfeccionamiento de la práctica psicológica en los distintos contextos.

REFERENCIAS

- Abal, F. (2013). *Comparación de modelos dicotómicos y politómicos de la Teoría de Respuesta al Ítem aplicados a un test de comportamiento típico*. [Tesis Doctoral, Universidad de Buenos Aires].
- Abal, F. J. P., Auné, S. E., & Attorresi, H. F. (2019). Construcción de un banco de ítems de facetas de neuroticismo para el desarrollo de un test adaptativo. <http://dx.doi.org/10.18682/pd.v1i1.854>.
- Abal, F. J. P., Lozzia, G. S., Aguerri, M. E., Galibert, M. S., & Attorresi, H. F. (2010). La escasa aplicación de la teoría de respuesta al ítem en tests de ejecución típica. *Revista Colombiana de Psicología*, *19*(1), 111-122.
- Abal, F. J. P., Lozzia, G., Aguerri, M. E., Galibert, M. S., & Attorresi, H. F. (2009). *Interpretación de los parámetros de modelos dicotómicos de la TRI aplicados a la medición de la personalidad*. Memorias Congreso Internacional de Investigación y Práctica Profesional en Psicología XVI. Facultad de Psicología, Universidad de Buenos Aires. <https://www.aacademica.org/000-020/743>.
- Abal, F. J. P., Lozzia, G., Picón Janeiro, J., Galibert, M. S., & Aguerri, M. E. (2007). *Dificultades en la difusión y el desarrollo de la teoría de respuesta al ítem en Argentina*. Memorias de XIV Jornadas de Investigación y Tercer Encuentro de Investigadores en Psicología del Mercosur. Facultad de Psicología, Universidad de Buenos Aires. <https://www.aacademica.org/000-073/588>.
- Abal, F. J. P., Sanchez Gonzalez, J. F. & Lozzia, G. S. (2020). *Administración online versus lápiz-y-papel: funcionamiento diferencial de ítems de Neuroticismo*. Memorias del VII congreso internacional de investigación en psicología. Facultad de psicología, Universidad Nacional de la Plata.
- Baker, F. B., & Kim, S. H. (2017). *The basics of item response theory using R* (pp. 55-67). Springer.
- Bech, P., Carrozzino, D., Austin, S. F., Møller, S. B., & Vassend, O. (2016). Measuring euthymia within the Neuroticism Scale from the NEO Personality Inventory: a Mokken analysis of the Norwegian general population study for scalability. *Journal of Affective Disorders*, *193*, 99-102. <https://doi.org/10.1016/j.jad.2015.12.039>.
- Bejar, I. I. (1977). An application of the continuous response level model to personality measurement. *Applied Psychological Measurement*, *1*(4), 509-521. <https://doi.org/10.1177/014662167700100407>
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*(1), 29-51.
- Brown, R. D., & Harvey, R. J. (2003). *Detecting personality test faking with appropriateness measurement: Fact or fantasy*. Annual Conference of the Society for Industrial and Organizational Psychology, Orlando.
- Carter, J. E., & Wilkinson, L. (1984). A latent trait analysis of the MMPI. *Multivariate Behavioral Research*, *19*(4), 385-407. <https://doi.org/10.1207/s15327906mbr19042>.
- Colledani, D., Anselmi, P., & Robusto, E. (2019). Development of a new abbreviated form of the Eysenck personality questionnaire-revised with multidimensional item response theory. *Personality and Individual Differences*, *149*, 108-117. <https://doi.org/10.1016/j.paid.2019.05.044>.
- Escorial, S., & Navas, M. J. (2007). Analysis of the gender variable in the Eysenck Personality Questionnaire-revised scales using differential item functioning techniques. *Educational and Psychological Measurement*, *67*(6), 990-1001. <https://doi.org/10.1177/0013164406299108>.
- Ferrando, P. J. (1994). Fitting Item Response Models to the Epi-A Impulsivity Sub Scale. *Educational and Psychological Measurement*, *54*(1), 118-127. <https://doi.org/10.1177/0013164494054001016>.
- Ferrando, P. J. (2008). The impact of social desirability bias on the EPQ-R item scores: An item response theory analysis. *Personality and Individual Differences*, *44*(8), 1784-1794. <https://doi.org/10.1016/j.paid.2008.02.005>.
- Fraley, R. C., Waller, N. G. & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology*, *78*, 350-365. <https://doi.org/10.1037/0022-3514.78.2.350>.
- García Cueto, E., Muñiz, J., & Lozano, L. M. (2003). *Efecto de la reducción de alternativas en las escalas tipo Likert desde la perspectiva de la TRI*. Memorias IX Conferencia Española de Biometría, La Coruña (pp. 28-30).
- Highhouse, S., Zickar, M. J., Brooks, M. E., Reeve, C. L., Sarkar-Barney, S. T., & Guion, R. M. (2016). A public-domain personality item bank for use with the Raymark, Schmit & Guion (1997) PPRF. *Personnel Assessment and Decisions*, *2*(1), 5. <https://doi.org/10.25035/pad.2016.005>.
- Johnson, W., Spinath, F., Krueger, R. F., Angleitner, A., & Riemann, R. (2008). Personality in Germany and Minnesota: An IRT-based comparison of MPQ self-reports. *Journal of personality*, *76*(3), 665-706. <https://doi.org/10.1111/j.1467-6494.2008.00500.x>.
- Kramp Denegri, U. A. (2006). *Efecto del número de opciones de respuesta sobre las propiedades psicométricas de los cuestionarios de personalidad*. [Tesis Doctoral, Universitat de Barcelona]. <https://hdl.handle.net/2445/42554>.
- Kuncel, R. B. (1973). Response processes and relative location of subject and item. *Educational and Psychological Measurement*, *33*(3), 545-563. <https://doi.org/10.1177/001316447303300302>
- Lawley, D. N. (1943). On problems connected with item selection and test construction. *Proceedings of Royal Society of Edimburg*, *62*, 74-82.
- Luo, H., Cai, Y., & Tu, D. (2020). Procedures to Develop a Computerized Adaptive Testing to Advance the Measurement of Narcissistic Personality. *Frontiers in psychology*, *11*, 1437. <https://doi.org/10.3389/fpsyg.2020.01437>.
- Luzardo, M. (2017). *Contribuciones a la Teoría de Respuesta al Ítem no Paramétrica*. [Tesis Doctoral, Universidad Autónoma de Madrid]. <https://repositorio.uam.es/bitstream/handle/10486/681395>.
- Maydeu-Olivares, A. (2005). Further empirical results on parametric versus non-parametric IRT modeling of Likert-type personality data. *Multivariate Behavioral Research*, *40*(2), 261-279. https://doi.org/10.1207/s15327906mbr4002_5.
- McBride, N. L. (2001). *An item response theory analysis of the scales from the international personality item pool and the neo personality inventory-revised* [Tesis de Maestría, Virginia Tech]. <http://hdl.handle.net/10919/34430>.

- Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2007). Are Internet and paper-and-pencil personality tests truly comparable? An experimental design measurement invariance study. *Organizational Research Methods, 10*(2), 322-345. <https://doi.org/10.1177/1094428106289393>.
- Mokken, R. J. (1971): *A theory and procedure of scale analysis*. De Gruyter Mouton.
- Morizot, J., Ainsworth, A. T. & Reise, S. P. (2007). Toward modern psychometrics. application of item response theory models in personality research. En R. W. Robins, R. C. Fraley & R. F. Krueger (eds.), *Handbook of Research Methods in Personality Psychology* (pp. 407-423). Guilford Press.
- Muñoz, J. (2018). *Introducción a la psicometría: Teoría Clásica y TRI*. Pirámide.
- Nieto, M. D., Abad, F. J., Hernández-Camacho, A., Garrido, L. E., Barrada, J. R., Aguado, D., & Olea, J. (2017). Calibrating a new item pool to adaptively assess the Big Five. *Psicothema, 29*(3), 390-395. <https://doi.org/10.7334/psicothema2016.391>.
- Perez Sutil, J. M., Sanchez Gonzalez, J. F., & Ursino, D. (2019). *Aplicación de teoría de respuesta al ítem a la dimensión extraversión del EPQ-R*. Memorias XI Congreso Internacional de Investigación y Práctica Profesional en Psicología. Facultad de Psicología, Universidad de Buenos Aires.
- Reise, S. P. & Waller, N. G. (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement, 14*, 45-58. <https://doi.org/10.1177/014662169001400105>.
- Richardson, M. W. (1936). The relationship between difficulty and the differential validity of a test. *Psychometrika, 1*, 33-49.
- Sheppard, R., Han, K., Colarelli, S. M., Dai, G., & King, D. W. (2006). Differential item functioning by sex and race in the Hogan Personality Inventory. *Assessment, 13*(4), 442-453. <https://doi.org/10.1177/1073191106289031>.
- Spence, R., Owens, M., & Goodyer, I. (2012). Item response theory and validity of the NEO-FFI in adolescents. *Personality and Individual Differences, 53*(6), 801-807. <https://doi.org/10.1016/j.paid.2012.06.002>.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2012). Constructing fake-resistant personality tests using item response theory: High-stakes personality testing with multidimensional pairwise preferences. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 214-239). Oxford University Press.
- Stepp, S. D., Yu, L., Miller, J. D., Hallquist, M. N., Trull, T. J., & Pilkonis, P. A. (2012). Integrating competing dimensional models of personality: Linking the SNAP, TCI, and NEO using Item Response Theory. *Personality Disorders: Theory, Research, and Treatment, 3*(2), 107-126. <https://doi.org/10.1037/a0025905>.
- Thurstone, L. L. (1927). A law comparative judgment. *Psychological Review, 24*, 273-286.
- Tucker, L. R. (1946). Maximum validity of a test with equivalent item. *Psychometrika, 11*, 1-13.
- Watson, R., Deary, I., & Austin, E. (2007). Are personality trait items reliably more or less 'difficult'? Mokken scaling of the NEO-FFI. *Personality and Individual Differences, 43*(6), 1460-1469. <https://doi.org/10.1016/j.paid.2007.04.023>.