

# Validación del “Toronto Empathy Questionnaire (TEQ)” en una Muestra Uruguaya

## Validation of the “Toronto Empathy Questionnaire (TEQ)” in a Uruguayan Sample

Marcela Carballo<sup>1</sup>, Cesar Daniel Costa-Ball<sup>2</sup>, Martín Bidegain<sup>3</sup> y José Álvarez<sup>4</sup>

### Resumen

La evaluación de la empatía se ha realizado a través de diferentes escalas de autorreporte que estudian aspectos afectivos o cognitivos del constructo. A pesar de la aceptación actual de las escalas multidimensionales, se ha sugerido que es posible evaluar una dimensión general subyacente, siendo el Toronto Empathy Questionnaire (TEQ) una técnica desarrollada para este propósito. Este estudio analizó las propiedades psicométricas del TEQ en una muestra de 200 adultos uruguayos (100 mujeres, 100 hombres). El análisis factorial mostró un buen ajuste para el modelo unidimensional. Eliminando 10 ítems se alcanza la unidimensionalidad esencial. Los resultados sugieren que la versión del TEQ en español es válida para la población en estudio. Se sugiere revisar la formulación de los ítems invertidos y la validez de la versión reducida de 6 ítems.

**Palabras clave:** empatía, TEQ, validación, propiedades psicométricas

### Abstract

Assessing empathy has been carried out through different self-report scales that study the affective or cognitive aspects of the construct. Despite the current acceptance of multidimensional scales, it has been suggested that it is possible to assess an underlying general dimension, the Toronto Empathy Questionnaire (TEQ), a technique developed for this purpose. This study analyzed the psychometric properties of the TEQ in a sample of 200 Uruguayan adults (100 women, 100 men). The exploratory factor analysis showed a good fit for the one-dimensional model. By eliminating 10 items, essential unidimensionality is reached. The results suggest that the Spanish version of the TEQ is a valid scale for the study population. A revision of the formulation of the inverted items and the validity of the reduced 6-item version is suggested.

**Keywords:** empathy, TEQ, validation, psychometric properties

<sup>1</sup>Doctora en Psicología. Departamento de Neurociencia y Aprendizaje, Universidad Católica del Uruguay, Comandante Braga 2715, Montevideo, CP 11600, Uruguay. Tel.: +59824872717. Correo: marcela.carballo@ucu.edu.uy

<sup>2</sup>Doctor en Psicología. Departamento de Neurociencia y Aprendizaje, Universidad Católica del Uruguay, Comandante Braga 2715, Montevideo, CP 11600, Uruguay. Tel.: +59824872717. Correo: ccosta@ucu.edu.uy

<sup>3</sup>Licenciado en Psicología. Departamento de Neurociencia y Aprendizaje, Universidad Católica del Uruguay, Comandante Braga 2715, Montevideo, CP 11600, Uruguay. Tel.: +59824872717. Correo: mignacio.bidegain@ucu.edu.uy

<sup>4</sup>Licenciado en Psicología. Departamento de Neurociencia y Aprendizaje, Universidad Católica del Uruguay, Comandante Braga 2715, Montevideo, CP 11600, Uruguay. Tel.: +59824872717. Correo: noel.alvarez@ucu.edu.uy

## Introducción

La empatía es una habilidad fundamental para la interacción social, la comunicación y el desarrollo de habilidades sociales. A pesar de su relevancia para el desarrollo y convivencia saludables, aún no existe consenso respecto a su definición, los mecanismos que la integran y la forma de evaluarla. Se trata de un constructo complejo, cuya conceptualización ha ido cambiando a lo largo de los años pasando por consideraciones más afectivas o cognitivas. Actualmente, han tomado mayor fuerza teorías que consideran que se trata de un fenómeno multidimensional, que involucra tanto aspectos afectivos como cognitivos (Decety & Jackson, 2004; Cuff et al., 2016), a lo cual se puede agregar un componente motivacional o conductual, definido como una tendencia a actuar en consonancia con la emoción percibida (Batchelder et al., 2017; Decety et al., 2016).

Las escalas de autorreporte han sido ampliamente usadas para el estudio de la empatía, evaluando principalmente aspectos afectivos y cognitivos del constructo, mientras que el componente conductual o motivacional ha sido más estudiado con métodos de observación comportamental (Arrieta, et al., 2021). Las escalas que evalúan distintos componentes de forma independiente son de gran aceptación en la actualidad. Entre ellas, el Índice de Reactividad Interpersonal (IRI), fue la escala más usada para investigar la empatía entre los años 2000 y 2015 (Ilgunaite et al., 2017; Hall & Schwartz, 2019) y evalúa discriminadamente aspectos cognitivos y afectivos. Otros instrumentos, como la Escala de Conductas Empáticas (ECE, Auné et al., 2017) ponen el foco en el componente conductual del constructo.

Sin embargo, a pesar de existir cierto consenso en considerar a la empatía como un constructo compuesto por múltiples componentes, la independencia de éstos se ha puesto en cuestionamiento. Algunos autores han planteado la existencia de una dimensión general en la disposición empática, que contiene a los diferentes componentes. Proponen una aproximación unidimensional, complementaria al estudio centrado en las diferentes dimensiones (Cliffordson, 2002; Spreng et al., 2009).

Desde esta perspectiva, el TEQ (Spreng et al., 2009), se desarrolla a partir del estudio de las principales escalas, con el objetivo de rescatar los aspectos comunes a través de ítems altamente correlacionados. Para su desarrollo, los autores utilizaron 95 ítems pertenecientes a nueve escalas de empatía: IRI (Davis, 1983), Hogan's Empathy Scale (Hogan, 1969), Questionnaire Measure of Emotional Empathy (Mehrabian & Epstein, 1972), Balanced Emotional Empathy Scale (Mehrabian, 2000), the Scale of Ethnocultural Empathy (Wang et al., 2003), Jefferson Scale of Physician Empathy (Hojat et al., 2001), Nursing Empathy Scale (Reynolds et al., 2000), Japanese Adolescent Empathy Scale (Hashimoto & Shiomi, 2002), y The Measure of Emotional Intelligence (Schutte et al., 1998). Además, se incluyeron 36 ítems nuevos, propuestos por los autores en base a la literatura sobre el tema. Los autores usando el análisis factorial exploratorio (AFE), obligaron al conjunto de ítems a cargar en un solo factor, utilizando una serie de tres estudios, para construir TEQ. Como resultado de este análisis surge una escala conformada por 16 ítems que evalúan diferentes atributos relacionados a la empatía. El aspecto más comúnmente evaluado en estas escalas y que da lugar al TEQ es un proceso principalmente emocional, vinculado a una captación afectiva del estado del otro. Esta dimensión incluye procesos descritos como contagio emocional, comprensión emocional, activación fisiológica de tipo simpática y altruismo.

Trabajos posteriores tradujeron la escala a diferentes idiomas y realizaron estudios de validación y análisis de propiedades psicométricas. Totan et al. (2012) en un estudio de validación de la versión turca, confirmaron el carácter unidimensional del instrumento. Sin embargo, se señaló que los ítems 1, 6 y 9 tenían bajo poder de predictibilidad y baja carga factorial, lo que llevó a excluirlos en esa versión. A su vez, encontraron correlaciones significativas entre el TEQ y otras escalas de empatía como la Empathic Tendency Scale (ETS) y el Basic Empathy Scale (BTS).

Kourmoussi et al. (2017) examinaron la validez y la confiabilidad del cuestionario, traducido al idioma griego, confirmando el modelo de un factor. Sin embargo, encontraron que la carga factorial del ítem 11 era baja y el modelo mostraba un mayor ajuste al excluir este ítem.

Estudios con la versión del instrumento en idioma chino (Xu et al., 2020) encontraron que la estructura que más ajustaba presentaba tres factores, los cuales denominaron: empatía positiva (ítems 1, 2, 3, 5, 6, 13, 16), empatía neutra (ítems 8 y 9) y empatía negativa (ítems 4, 7, 10, 11, 12, 14, 15). A su vez, consideran necesario el análisis de los ítems 7 y 14 ya que no proporcionan buenos valores test-retest y su exclusión mejora los indicadores.

Manavipour y Memarian (2020) estudiaron las propiedades psicométricas en una muestra de estudiantes universitarios iraníes y proponen una estructura con un factor general, de primer orden, y cinco factores específicos: indiferencia (ítems 2, 4, 12, 14, 15), empatía reactiva (ítems 8, 1, 9), impresión emocional (ítems 7, 10), estabilidad emocional (ítems 3, 5, 16) y empatía emocional (ítems 6, 13).

Kim y Han (2016) a partir de dos estudios con la versión coreana encontraron una adecuada confiabilidad test-retest y consistencia interna. Si bien el AFE sugería un modelo con un único factor, el AFC presentó fallas para confirmarlo. Yeo y Kim (2021), en otra investigación con estudiantes de medicina coreanos, observaron que la versión de 16 ítems tenía un poder explicativo débil. Los resultados sugieren adoptar un modelo de tres factores con 14 ítems (excluyendo los ítems 3 y 4).

Novak et al. (2021) evaluaron la estructura factorial y validez del TEQ en población checa, Tanto el AFE como el AFC apoyan la solución unidimensional observándose una consistencia interna elevada. Al mismo tiempo se señala que los ítems redactados negativamente pueden interferir con las propiedades psicométricas del cuestionario.

Si bien existen estudios que confirman la validez y confiabilidad de este instrumento para el estudio de la empatía en diferentes idiomas, estos aún son escasos y los resultados difieren respecto a la estructura de un factor y la adecuación de los ítems. No se cuenta con datos sobre la versión en español de este instrumento. Este estudio tiene como objetivo analizar la validez de contenido, constructo y concurrente en la versión del TEQ (Spreng et al., 2009) traducida al idioma español, en una muestra de adultos uruguayos.

## Método

### Participantes

La muestra estuvo compuesta por 200 participantes uruguayos, 100 hombres y 100 mujeres, con edades comprendidas entre los 18 y los 74 años ( $M=36.8$ ;  $DE=15.15$ ) y nivel socioeconómico (NSE) medio y alto. El 33.5% corresponden a un NSE medio (Medio bajo=4.0; Medio=9.0; Medio alto = 20.5) y el 66.5% restante pertenece al NSE alto (Alto=27.5; Muy alto=39.0). El 99% de los participantes ha completado la educación primaria, el 5.5% cuenta con estudios de secundaria incompletos, el 11.5% culminó secundaria o escuela técnica, el 40.5% posee estudios terciarios incompletos, y el 41.5% completó estudios terciarios.

### Instrumentos

Toronto Empathy Questionnaire (TEQ) (Spreng et al., 2009)<sup>1</sup>. Se trata de un cuestionario de autorreporte, diseñado para la evaluación de la empatía. Consta de 16 ítems, 8 se puntúan formulados negativamente y 8 positivamente. Se responde con una escala tipo Likert de cinco puntos (0=nunca; 1=casi nunca; 2=algunas veces; 3=casi siempre; 4=siempre), según la frecuencia con que la persona siente o actúa de esa manera. Los resultados se obtienen a partir de la suma de las respuestas, conformando una puntuación global del nivel de empatía. Las propiedades psicométricas del TEQ (Spreng et al., 2009) presentan un ajuste a un modelo unidimensional a través de una estructura factorial única compuesta por 16 ítems, con carga factorial superior a .40, presentando una alta fiabilidad reportada con el valor  $\alpha$  de Cronbach de .85, correlacionando en forma positiva y fuerte con las subescalas preocupación empática y toma de perspectiva de la escala IRI de Davis (1980, 1983).

Índice de Reactividad Interpersonal (IRI) (Davis, 1983): es una escala de autorreporte utilizada para el estudio de la empatía, basada en una perspectiva multidimensional. Está compuesta por cuatro subescalas que evalúan, de forma independiente, aspectos cognitivos y afectivos del constructo. Las escalas Fantasía (F) y Toma de

<sup>1</sup> Se utilizó una versión del cuestionario traducido al español por el equipo de LIF-UC (<https://www.liflab.cl/>)

Perspectiva (TP), evalúan componentes cognitivos y las escalas Preocupación Empática (PE) y Malestar Personal (MP) que evalúan aspectos afectivos. Consta de 28 ítems, siete para cada subescala, que se responden con una escala tipo Likert de cinco puntos indicados por letras A (no me describe bien), B, C, D, E (me describe muy bien). Los ítems contienen afirmaciones sobre sentimientos y el participante debe responder en qué medida considera que esa afirmación lo describe. Los valores de fiabilidad del estudio original de Davis (1983) fueron adecuados para las cuatro subescalas ( $F=.75$ ,  $TP=.75$ ,  $PE=.72$  y  $MP=.78$ ). En este estudio se utilizó la versión adaptada a población chilena por Fernández et al. (2011), con similares niveles de fiabilidad ( $F=.76$ ,  $EC$  y  $TP=.73$ ,  $MP=.70$ ).

Índice de Nivel Socioeconómico (INSE): es un cuestionario desarrollado en Uruguay para evaluar el nivel socioeconómico de las familias. Se utilizó la versión reducida formada por seis ítems que permite realizar una clasificación de los hogares en base a su nivel socioeconómico, infiriendo la capacidad de consumo a partir de información sociodemográfica y posesión de activos tanto tangibles como intangibles (Llambí & Piñeyro, 2012).

## Procedimiento

Se llevó a cabo un estudio instrumental (Montero & León, 2007) utilizando la Teoría Clásica de los Tests (Muñiz, 2010), con el propósito de realizar la adaptación transcultural del instrumento TEQ al Uruguay. Se utilizó la traducción del TEQ adaptada para Chile, y se pasó a la etapa de estudio de validez de contenido, basado en juicio de expertos, utilizando el método de coeficiente de validez de contenido (CVC) propuesto por Hernández-Nieto (2002). Este método evalúa la claridad del lenguaje, coherencia, relevancia y suficiencia, conceptos propuestos por Escobar-Pérez y Cuervo-Martínez (2008). Este coeficiente mide el grado de concordancia entre los jueces respecto a cada ítem, así como para el instrumento en su conjunto. Se puso el total de ítems (16) a consideración de tres jueces expertos, quienes fueron contactados por correo electrónico solicitando responder a un cuestionario autoadministrado para evaluar el contenido y la relevancia de los ítems en relación con el dominio

teórico que evalúan. Los expertos evaluaron las cuatro dimensiones. En el caso de la suficiencia se realiza una única evaluación considerando todos los ítems que conforman la escala, las tres restantes dimensiones se evalúan en cada ítem. Se utilizó una puntuación del 1 al 4 para evaluar cada ítem (1=no cumple con el criterio, 2=bajo nivel, 3=moderado nivel, 4=alto nivel).

La muestra fue obtenida a través de un muestreo no probabilístico, con la técnica bola de nieve, por difusión del estudio a través de redes sociales e invitando a participar a estudiantes universitarios. Luego de acceder y aceptar el consentimiento informado, los participantes completaron las escalas TEQ, IRI e INSE, utilizando la plataforma Qualtrics (Qualtrics, 2021). El procedimiento fue aprobado por el Comité de Ética de la Universidad Católica del Uruguay.

El estudio psicométrico comenzó por el análisis de los estadísticos descriptivos y prueba de normalidad para cada ítem. La validez de constructo se estudió a través de un Análisis Factorial Exploratorio (AFE) y se confirmó con el método de unidimensionalidad esencial, ambos implementados con el programa FACTOR (Lorenzo-Seva & Ferrando, 2006; Ferrando & Lorenzo-Seva, 2017). A continuación, se calculó la fiabilidad y finalmente la validez concurrente, correlacionando la escala TEQ con los resultados en las cuatro dimensiones del IRI (Davis, 1980, 1983).

## Análisis de Datos

Para el estudio de la validez de contenido del instrumento se empleó el método de Coeficiente de Validez de Contenidos CVC (Hernández-Nieto, 2002). Se usó el criterio menos estricto para conservar los ítems: valores de  $CVC > .70$ . (Balbinotti, 2004). Hernández-Nieto (2002) recomienda  $CVC > .80$ , valores  $.70 > CVC < .80$  se consideran valores dudosos y valores  $< .70$  insatisfactorios (Balbinotti et al., 2007).

Para la realización del Análisis Factorial Exploratorio y analizar sus propiedades psicométricas se siguió las recomendaciones y procedimientos actuales para su implementación (Ledesma et al., 2019; Lloret-Segura et al., 2014; Lloret et al., 2017).

Se calcularon los estadísticos descriptivos de los ítems (media, desviación típica, asimetría, curtosis, y prueba de normalidad), así como el índice de

homogeneidad corregido. Se realizó el análisis de ítems y se puso a prueba la suposición de normalidad mediante el estadístico de Kolmogorov-Smirnov, y las pruebas multivariadas de asimetría y curtosis propuestas por Mardia (1970). Distribuciones univariadas con coeficientes de asimetría mayores en valor absoluto de 1 (Ferrando & Anguiano-Carrasco, 2010; Flora et al., 2021) y valores de curtosis  $>8$  (Kline, 2005) se consideran evidencia de problemas de normalidad multivariada, que se confirman cuando el estadístico de Mardia presenta un nivel crítico  $p < .05$  y llevan al rechazo del supuesto de la distribución de normalidad en asimetría y curtosis.

El AFE se llevó a cabo con el programa FACTOR. Se realizó la prueba de adecuación de la matriz de correlación mediante la prueba de esfericidad de Barlett y la medida de Kaiser-Meyer-Olkin. La adecuación de los datos al AFE se considera “suficiente” con valores  $KMO > .70$  y “satisfactoria” con valores  $> .80$  (Lloret-Segura et al., 2014). Con muestras pequeñas ( $N=200$  casos), con ítems ordinales con cinco o más categorías de respuestas, con distribuciones aproximadamente normales, o coeficientes de asimetría entre valores  $-2$  y  $2$ , con coeficientes de curtosis  $<7$  (Bandalos & Finney, 2018) o  $<8$  (Kline, 2005), es aconsejable usar la matriz de correlaciones de Pearson, utilizando el método de estimación ML: Robust Maximum Likelihood (Lloret-Segura et al., 2014).

Para establecer el número de factores a retener se utilizaron dos métodos implementados en el programa FACTOR: Análisis Paralelo (PA, Timmerman & Lorenzo-Seva, 2011), y el método Hull (Lorenzo-Seva et al., 2011). Además, a nivel de ítem, se utilizaron los índices de evaluación general para determinar la unidimensionalidad esencial implementada en el programa FACTOR, para discriminar si con la totalidad de los ítems el modelo responde a una lógica unidimensional o multidimensional (Ondé & Alvarado, 2022): a) Congruencia unidimensional (UniCo), b) Varianza común explicada (VCE) y, c) Media de cargas absolutas residuales de ítem (MIREAL). Los datos se pueden tratar como esencialmente unidimensionales cuando UniCo  $> .95$ , ECV  $> .85$  y MIREAL  $> .30$  (Ferrando & Lorenzo-Seva, 2017), siendo recomendable que las saturaciones o carga factorial  $> .40$  (Lloret-Segura et al., 2014). Se realizó la depuración de la escala de 16 ítems, eliminando ítems

con comunalidad baja, y a continuación se ejecutó nuevamente el AFE. El procedimiento se repitió hasta obtener índices de bondad de ajuste e índices de unidimensionalidad esencial que ajustaran al modelo unidimensional esencial.

Para evaluar el ajuste del modelo, se utilizó el índice de ajuste de chi-cuadrado. Un valor bajo y un valor  $p$  no significativo ( $p > .05$ ) indican un buen ajuste, que implica que no hay diferencia significativa entre las matrices de covarianza observadas y estimadas. Por otro lado, existen limitaciones en el uso del estadístico chi-cuadrado como índice de ajuste del modelo, al ser un índice sensible al tamaño de la muestra. Con tamaños muestrales grandes ( $N > 100$ ), disminuye el valor  $p$  y tiende a ser significativo (Marsh et al., 1988). Para corregir este problema, se utilizó el chi-cuadrado/ $gl < 2$  o 3 (Schreiber et al., 2006) y dos rutas posibles: Comparative Fit Index (CFI)  $> .95$ ; Tucker Lewis Index (TLI)  $> .95$  y Root Mean Square Error of Approximation (RMSEA)  $\leq .07$  o CFI  $> .95$  y Standardized Root Mean Residual (SRMR)  $< .08$ , para escalas entre 12 y 30 ítems y con tamaño muestrales  $N < 250$  (Hair et al., 2014).

Para examinar la consistencia interna de la escala se utilizaron dos métodos implementados en el programa FACTOR: el estadístico omega de McDonald (Ventura-León & Caycho-Rodríguez, 2017) y el procedimiento del coeficiente alfa ordinal (Elosua-Oliden & Zumbo, 2008). Estos dos métodos son adecuados a la naturaleza ordinal de las variables estudiadas y a la falta de normalidad multivariante (Elosua-Oliden & Zumbo, 2008). Se consideró el valor  $\geq .70$  como una fiabilidad adecuada para su uso en ámbitos de investigación (Hair et al., 2014) y fiabilidad mínima de  $\geq .80$  para su uso en la clínica (Pérez & Carretero-Dios, 2005). Por último, se sugiere que la decisión de eliminar o conservar un ítem sea tomada considerando los índices estadísticos en función a la delimitación conceptual y la validez de contenido (Pérez & Carretero-Dios, 2005).

Se obtuvo la validez convergente a partir de las correlaciones entre el resultado del TEQ y las cuatro dimensiones de empatía evaluadas en la escala IRI. Se utilizó el coeficiente de Correlación de Pearson ( $r$ ). Los valores de correlación expresan:  $r \geq .20$  correlación baja,  $r \geq .50$  moderada,  $r \geq .80$  correlación fuerte (Ferguson, 2016).

Tabla 1. Estudio de Validez de Contenido de la Escala TEQ: Acuerdo entre Jueces

	Coh	Rel	Cla
1. Cuando otras personas demuestran entusiasmo, tiendo a sentirme entusiasmado también.	.96	.96	.96
2. Los problemas de otras personas me afectan poco. (I) <sup>a</sup>	.96	.88	.96
3. Me molesta que alguien sea tratado de manera poco respetuosa.	.96	.88	.96
4. Cuando alguien cercano a mí está contento. mi actitud no cambia. (I)	.96	.96	.96
5. Me gusta cuando consigo que otras personas se sientan mejor.	.88	.80	.96
6. Tengo una actitud cariñosa y considerada hacia personas que tienen menos suerte que yo	.96	.88	.96
7. Cuando un(a) amigo(a) empieza a contarme sus problemas trato de cambiar de tema (I)	.96	.96	.96
8. Soy capaz de darme cuenta cuando los demás están tristes incluso si no me dicen nada al respecto	.96	.96	.96
9. En general. "sintonizo" con los estados de ánimo de los demás.	.96	.96	.96
10. No soy comprensivo con la gente que se causa a sí misma enfermedades severas (I)	.88	.80	.88
11. Me molesta cuando alguien llora (I)	.96	.96	.96
12. En realidad no me interesa mucho cómo se sienten los demás (I)	.96	.88	.96
13. Cuando veo que otra persona está muy preocupada. siento una necesidad urgente de ayudar.	.96	.88	.96
14. Cuando veo que una persona es tratada de manera injusta. no siento mucha lástima por ella. (I)	.96	.88	.96
15. Encuentro absurdo que la gente ande buscando desesperadamente la felicidad (I)	.88	.71	.96
16. Cuando veo que se están aprovechando de alguien me dan ganas de protegerlo.	.88	.88	.96

Nota. Coh=coherencia; Rel=relevancia; Cla=claridad

<sup>a</sup> (I)=ítems invertidos

## Resultados

### Análisis de Validez de Contenido con procedimiento de Acuerdo entre Jueces

Para la validación de contenido, los ítems de la escala de empatía (TEQ) fueron analizados por tres especialistas con vasta experiencia en el área de los procesos afectivos: un psicólogo estudiante de doctorado en el área de la regulación afectiva, una doctora en psicología especialista en regulación emocional, y una psicóloga, doctora en biología y experta en investigación en empatía. En primer lugar, se evaluó la suficiencia del conjunto de ítems que conforman la escala TEQ para medir la empatía, en segundo lugar, se evaluó cada ítem, en relación con la claridad de lenguaje, coherencia teórica con el constructo y relevancia teórica. Se utilizó el coeficiente de validez de contenido (CVC) de Hernández-Nieto (2002). En la Tabla 1 se muestran los respectivos valores del coeficiente CVC en cada una de las dimensiones evaluadas en tres de las cuatro dimensiones evaluadas (coherencia, relevancia y claridad), obteniéndose en 15 de los 16 ítems valores de  $CVC > .80$  que evidencian validez y concordancias muy satisfactorias. Únicamente el ítem 15 para la dimensión relevancia presentó un valor  $CVC = .71$ , siendo esta aceptable o susceptible de revisar. En la dimensión suficiencia, para el total de los 16 ítems de la escala, se obtuvo un valor de  $CVC = .96$ .

### Análisis descriptivo de los ítems y validez de constructo

En la Tabla 2 se muestran los estadísticos descriptivos y prueba de normalidad mediante el estadístico de Kolmogorov-Smirnov de los 16 ítems de la escala de empatía (TEQ). Del total de ítems, tres presentaron desvío típico reducido (ítems 3, 5 y 7). En todos los ítems el coeficiente de asimetría es negativo y osciló entre  $-.20$  y  $-.204$ . Seis de 16 ítems presentaron coeficientes de asimetría mayores a 1 en valor absoluto (Ferrando & Anguiano-Carrasco, 2010), evidenciando problemas de normalidad multivariada. Se rechaza la hipótesis de normalidad pudiéndose concluir que los ítems del TEQ no se ajustan a una distribución normal. Se utilizó el análisis de Mardia (1970) para la asimetría y curtosis multivariante, hallando un coeficiente de asimetría de  $57.36$ ,  $gl=816$ ,  $p=1.00$  y un coeficiente de curtosis de  $338.65$ ,  $p < .000$ , evidenciando la ausencia de una distribución de normalidad multivariante en los datos.

El análisis factorial se realizó con el programa FACTOR. Se decidió factorizar la matriz de datos, usando la matriz de correlaciones de Pearson, dada la naturaleza ordinal de los ítems que presentan cinco categorías de respuestas, con distribuciones de coeficientes de asimetría que oscilan entre  $-2$ ,  $2$ , con un tamaño de muestra ( $N=200$ ) con el cual no es recomendable el uso de las matrices policóricas al ser poco estables con muestras menores a 300 sujetos (Lloret-Segura et al., 2014; Lloret et al., 2017). Se utilizó el método de estimación de Máxima Verosimilitud (RML: *Robust Maximum*

Tabla 2. Descriptivos Univariados de la Escala de Empatía de Toronto (TEQ)

Ítems	M	DE	Asimetría	Curtosis	K-S	Ihc
1	3.65	.79	-.82	2.60	.27***	.46
2	3.37	.89	-.20	.05	.23***	.17
3	4.53	.63	-1.03	-.02	.38***	.60
4	3.59	1.14	-.58	-.45	.24***	-.02
5	4.73	.53	-2.04	4.38	.46***	1.12
6	3.81	.94	-.63	.21	.24***	.36
7	4.56	.65	-1.76	4.47	.38***	.51
8	3.78	.79	-.28	.03	.26***	.39
9	3.48	.87	-.26	.18	.22***	.46
10	3.67	1.07	-.48	-.49	.22***	.12
11	4.05	1.05	-1.05	.50	.24***	.12
12	4.32	.88	-1.20	.79	.32***	.55
13	3.76	.78	-.45	.62	.28***	.73
14	4.38	.94	-1.85	3.22	.39***	.23
15	3.77	1.16	-.53	-.60	.23***	.06
16	4.14	.76	-.59	-.10	.24***	.60

Nota. K-S=prueba de Kolmogorov-Smirnov; Ihc=índice de los ítems corregidos.

\*\*\* $p < .000$

Tabla 3. Indicadores de bondad de ajuste para la unidimensionalidad esencial

Ítems	Excluir	$\chi^2$	gl	P	CFI	TLI	RMSEA	RMSR	UniCo	ECV	I-Real
16	-	157.7	104	.00	.95	.94	.054	.075	.86	.77	.23
15	ítem 4	141.5	90	.00	.95	.94	.046	.072	.90	.78	.22
14	ítem 15	108.5	77	.01	.96	.96	.045	.072	.91	.78	.21
13	ítem 10	94.7	65	.01	.97	.96	.048	.072	.91	.79	.23
12	ítem 2	85.5	54	.01	.96	.95	.055	.073	.90	.78	.23
11	ítem 11	66.3	44	.02	.97	.96	.051	.068	.92	.80	.23
10	ítem 7	48.4	35	.06	.98	.97	.044	.066	.94	.79	.21
9	ítem 14	32.7	27	.20	.99	.98	.033	.056	.95	.79	.22
8	ítem 1	27.2	20	.13	.99	.98	.043	.056	.96	.81	.20
7	ítem 3	19.5	14	.15	.99	.98	.045	.058	.95	.81	.24
6	ítem 12	16.5	9	.06	.98	.97	.065	.058	.97	.85	.23

Nota.  $\chi^2$ =ajuste absoluto; gl=grados de libertad;  $p$ =nivel crítico; CFI=índice de ajuste comparativo; TLI=índice de Tucker-Lewis; RMSEA=error de aproximación cuadrático medio; RMSR=raíz cuadrada de la media de residuos cuadrados; Uni\_Co=índice de congruencia de la matriz de carga factorial y la matriz de carga esperada; ECV=varianza común explicada; I-REAL=magnitud de las cargas correspondientes al segundo factor canónico.

*Likelihood*). La adecuación de los datos al AFE arrojó resultados adecuados, las medidas KMO=.81 y el estadístico de Bartlett=606.2 (gl=120,  $p < .000$ ) indicaron que la matriz es apta para la factorización. El procedimiento de análisis paralelo, el método HULL y BIC, recomiendan la extracción de un único factor, que explicó el 26% de la varianza total. Los índices de bondad de ajuste del AFE de la solución factorial muestran un buen ajuste a un modelo unidimensional (chi-cuadrado=157.7; gl=104;  $p=.000$ ; chi-cuadrado/gl=1.51; RMSEA=.051; RMSR=.075; CFI=.95; TLI=.94) aunque no ajusta a un modelo de unidimensionalidad esencial, con UniCo de 0.86, ECV de .77 y MIREAL de .23, dos de los tres índices están por fuera de los límites habituales (.95, .85 y .30). Todas las cargas factoriales oscilaron entre .08 y .75, nueve de los 16 ítems presentaron cargas superiores a .40. La consistencia del modelo

unidimensional fue adecuada con un alfa ordinal de .77.

En la Tabla 3, se muestran los indicadores de bondad de ajuste obtenidos en cada paso del procedimiento para el análisis de la unidimensionalidad esencial. Con este procedimiento, se eliminaron 10 de los 16 ítems de la escala original (1, 2, 3, 4, 7, 10, 11, 12, 14, 15), obteniendo una escala abreviada de 6 ítems.

Los índices de bondad de ajuste del AFE de la solución factorial muestran un excelente ajuste a un modelo unidimensional esencial (chi-cuadrado=157.7, gl=104,  $p=.000$ , Chi cuadrado/gl=1.51; RMSEA=.051; RMSR=.075; CFI=.95; TLI=.94; UniCo de .97, ECV de .85 y MIREAL de .23, con los tres índices dentro de los límites habituales: .95, .85 y .30). Estos tres indicadores, colaboran en decidir si existe un factor fuerte y dominante para que el conjunto de datos

Tabla 4. Puntuaciones factoriales y comunalidad de AF con la escala total y la abreviada

	Escala completa 16 ítems		Escala abreviada 6 ítems	
	FI	Comunalidad	FI	Comunalidad
Ítem 1	.39	.16	-	-
Ítem 2	.27	.07	-	-
Ítem 3	.45	.20	-	-
Ítem 4	.08	.00	-	-
Ítem 5	.56	.31	.52	.27
Ítem 6	.58	.34	.60	.36
Ítem 7	.33	.11	-	-
Ítem 8	.50	.24	.54	.29
Ítem 9	.52	.26	.49	.24
Ítem 10	.22	.04	-	-
Ítem 11	.30	.08	-	-
Ítem 12	.52	.27	-	-
Ítem 13	.75	.55	.81	.66
Ítem 14	.35	.12	-	-
Ítem 15	.21	.04	-	-
Ítem 16	.63	.40	.64	.41
Fiabilidad: Alfa ordinal		.77		.77
Fiabilidad: McDonald's Omega		.82		.90
EAP Reliability		.83		.81
Factor Determinacy Index (FDI)		.91		.90
AP marginal reliability		.83		.81
Sensitivity ratio (SR)		2.22		2.06
Expected percentage of true differences (EPTD)		90%		89%
Media		63.66		23.71
Desvío Estándar (DE)		6.57		3.21
Asimetría		-.60		-.93
Curtosis		.40		1.86
Prueba de Normalidad: Kolmogorov-Smirnova		.08*		.11*

Nota. FI=carga factorial.

\* $p < .0$

Tabla 5. Estudio de validez convergente con las escalas IRI

	TP	FA	PE	MP
Escala 16 ítems (TEQ)	.39**	.24**	.67**	.07
Escala abreviada (TEQ)	.40**	.17*	.65**	.06

Nota. TP=toma de perspectiva; FA=fantasía; PE=preocupación empática; MP=malestar personal. Se presenta el nivel de correlación a partir del valor "r".

\*La correlación es significativa en el nivel .05 (bilateral). \*\* La correlación es significativa en el nivel .01 (bilateral).

pueda considerarse esencialmente unidimensional. Todas las cargas factoriales oscilaron entre .49 y .81. La consistencia del modelo unidimensional fue adecuada con el resultado obtenido con el alfa ordinal de .77, y satisfactoria con el coeficiente McDonald's Omega=.90.

La Tabla 4 muestra la matriz de carga factorial y las comunalidades en ambas soluciones factoriales. En la salida factorial con la escala completa, 10 de los 16 ítems que conforman la escala, presentaron carga factorial  $\geq .40$ , un ítem  $\geq .30$  y cinco ítems  $\leq .28$ , mientras que la salida con la escala abreviada con unidimensionalidad esencial, los 6 ítems presentan una carga factorial  $\geq .40$ .

### Análisis de Validez Convergente

En la Tabla 5 se muestra el análisis de la validez convergente de la escala de empatía TEQ, en su versión original y en la versión abreviada de 6 ítems, correlacionando los resultados con las 4 subescalas del IRI.

La evidencia de validez basada en la relación con otras variables se evaluó a partir del cálculo de las correlaciones del puntaje del TEQ con las 4 medidas de empatía de la escala Interpersonal Reactivity Index (IRI: Davis, 1983). Tanto para la escala completa como para la abreviada se encontraron correlaciones positivas con 3 de las 4 subescalas del IRI: Preocupación Empática, Toma de Perspectiva y Fantasía.

## Discusión

Este estudio tuvo como objetivo realizar una validación y análisis de propiedades psicométricas del TEQ en población uruguaya. Se trata de un instrumento desarrollado recientemente, que aparece como alternativa a la evaluación de la empatía desde una perspectiva multidimensional (Davis, 1980). Hasta el momento, únicamente se conocen adaptaciones en China (Xu et al., 2020), Corea (Kim & Han, 2016; Yeo & Kim, 2021), Grecia (Kourmoussi et al., 2017), Irán (Manavipour & Memarian, 2020) y República Checa (Novak et al., 2021). En Uruguay no existe adaptación del TEQ, así como en otros países de habla hispana. Este es el primer estudio de validación de la escala en idioma español y es, a su vez, la primera validación de una escala de empatía en población uruguaya.

En primer lugar, se llevó a cabo un estudio de validez de contenido con el procedimiento de coeficiente de validez de contenido (CVC) de Hernández-Nieto (2002). El análisis de la evaluación de los tres jueces expertos sobre la suficiencia, coherencia, relevancia y claridad de los ítems que componen la escala, mostró valores satisfactorios y aceptables de validez y concordancia para los 16 ítems. Estos resultados confirman que el TEQ es una escala válida para su aplicación en esta población desde el punto de vista del contenido de los ítems.

En segundo lugar, se llevó a cabo un análisis factorial que mostró un buen ajuste para el modelo de un factor para la escala completa de 16 ítems, cumpliendo los criterios de unidimensionalidad. Este resultado replica los hallazgos del estudio original de la escala en población canadiense (Spreng et al., 2009), que fueron posteriormente replicados en población turca (Totan, 2012), checa (Novak et al., 2021) y en docentes griegos (Kourmoussi, 2017).

En un análisis posterior se eliminaron los ítems que presentaban baja comunalidad hasta llegar a un modelo de unidimensionalidad esencial. Para esto se eliminaron los ítems 1, 2, 3, 4, 7, 10, 11, 12, 14, 15, dando lugar a una escala reducida conformada por los ítems 3, 5, 6, 8, 13, 16. Esta versión abreviada de 6 ítems cuenta con buenos indicadores de ajuste a la unidimensionalidad esencial, fiabilidad y validez concurrente alta con

la medida de preocupación empática de la escala IRI. Estos resultados sugieren la posibilidad de usar una escala reducida que permitiría evaluar el constructo utilizando únicamente 6 ítems, con la misma confiabilidad que la escala completa. Estos ítems incluyen la evaluación de aspectos que pueden ser entendidos como afectivos (*“Me gusta cuando consigo que otras personas se sientan bien”*, *“En general, “sintonizo” con los estados de ánimo de los demás”*), cognitivos (*“soy capaz de darme cuenta cuando los demás están tristes incluso si no me dicen nada al respecto”*), y conductuales o motivacionales (*“Tengo una actitud considerada y cariñosa hacia personas que tienen menos suerte que yo”*, *“Cuando veo que una persona está muy preocupada, siento una necesidad urgente de ayudar”*, *“Cuando veo que se están aprovechando de alguien, me dan ganas de protegerlo”*) reflejando, de forma más equilibrada, los aspectos considerados más relevantes en las conceptualizaciones actuales de la empatía (Arrieta et al., 2021).

Por otro lado, resulta importante hacer un análisis de los ítems eliminados en este proceso y considerar posibles problemas derivados de los mismos. De los 10 ítems que fueron descartados para alcanzar la unidimensionalidad esencial, se encuentran los ocho ítems redactados en sentido inverso, además de los ítems 1 y 3 redactados en sentido directo. La inclusión de ítems redactados tanto en forma directa como invertida es una estrategia generalizada, que ha sido recomendada en la creación de escalas (Abad et al., 2011), como forma de reducir el sesgo del estilo de respuesta (Nunnally, 1978). Sin embargo, el uso de formulaciones negativas de los ítems ha sido desaconsejado (Haladyna et al., 2002; Haladyna & Rodríguez, 2013) ya que puede afectar la fiabilidad y validez del instrumento al colaborar con la aparición de factores espurios que provocan el rechazo de modelos unidimensionales, a favor de modelos multidimensionales, que no presentan relación sustentable con la delimitación conceptual de los constructos a evaluar (Woods, 2006). La estrategia de utilizar ítems formulados positiva y negativamente, como en el caso del TEQ, se asocian a peores propiedades psicométricas y peor precisión de medición del instrumento, así como a problemas en la interpretación de la unidimensionalidad. Además, las respuestas se ven

más influenciadas por las habilidades verbales de los examinados (Suárez-Álvarez et al., 2018). Las publicaciones recientes sobre directrices para la construcción de instrumentos de evaluación psicológica sugieren tener cuidado con la redacción de los ítems negativos (Muñiz & Fonseca-Pedrero, 2019).

Un estudio reciente del TEQ con población checa (Novak et al., 2021), encontró que la reformulación de los ítems invertidos de la escala permitía obtener mejores parámetros psicométricos que la versión original con ítems invertidos. Los resultados del presente estudio apuntan en la misma dirección sugiriendo que es importante atender a los ítems de redacción invertida de la escala que pueden estar generando problemas. Futuros estudios deberían reformular estos ítems en la versión en español y estudiar las propiedades psicométricas de una versión compuesta únicamente por ítems redactados en forma directa.

Para analizar la validez convergente de la escala, se correlacionaron los datos del TEQ con los resultados obtenidos en la escala IRI. Esta escala es una de las más usadas para el estudio de la empatía en los últimos años (Ilgunaite et al., 2017; Hall & Schwartz, 2019). Se trata de un instrumento que evalúa el constructo desde una perspectiva multidimensional estudiando aspectos cognitivos y afectivos a través de cuatro subescalas. Las subescalas Preocupación Empática (PE) y Malestar Personal (MP), evalúan la dimensión afectiva, mientras que la dimensión cognitiva se evalúa con las subescalas Toma de Perspectiva (TP) y Fantasía (F).

Tanto para la escala completa de 16 ítems como para la versión abreviada del TEQ, los resultados de la correlación con las cuatro subescalas del IRI están en la línea de lo encontrado en el estudio original del TEQ (Spreng et al., 2009). En ambos estudios se encontró una correlación positiva fuerte con la escala PE, y una correlación moderada con las escalas TP y F, mientras que no se encontraron correlaciones con la subescala MP. El hecho de que la mayor correlación se encuentre con la escala PE es un resultado esperado considerando que el TEQ es una escala que parte de una conceptualización de la empatía como un proceso principalmente emocional (Spreng et al., 2009). Por otro lado, la ausencia de correlación con la otra subescala afectiva (MP) puede ser entendida

por el hecho, señalado por otros autores, de que esta subescala podría no estar evaluando aspectos centrales de la empatía, sino aspectos que tendrían más que ver con el autocontrol emocional o neuroticismo (Cliffordson, 2001; Baron-Cohen & Wheelwright, 2004; Baldner & McGinley, 2014; Spreng et al., 2009).

Considerando estos antecedentes, la correlación encontrada en este estudio con las subescalas del IRI (Davis, 1983), aportan evidencia a favor de la validez convergente. La replicación de los resultados del estudio original podría apoyar la idea de que las diferentes facetas de la empatía confluyen en una dimensión general, que puede ser medida con un instrumento unidimensional como el TEQ.

En relación con las limitaciones de la investigación, hay que destacar el bajo número de participantes ( $N = 200$ ), inferior al que en un principio se estimó alcanzar, teniendo en cuenta que para la utilización de procedimientos AFE o AFC, para datos ordinales o categóricos, se recomienda trabajar con muestras mayores a 400 participantes (Goretzko et al., 2021). El principal motivo fue el reducido número de participantes hombres que completaron el estudio. Con el objetivo de obtener una muestra equilibrada por la variable género, se aceptaron todos los protocolos completados por hombres y se eligió al azar un igual número de protocolos completados por mujeres. Otra limitación es referida a la baja representación en la muestra del NSE bajo. Se necesita buscar nuevas estrategias de muestreo para obtener mayor participación en el NSE bajo que por la estrategia usada no fue posible captar.

Por último, en relación con la validación del instrumento, es necesario volver a la etapa de construcción y evaluación cualitativa de ítems, con el objetivo de redactar los ítems invertidos como ítems directos, siguiendo la recomendación de las directrices para la construcción de instrumentos de evaluación psicológica (Haladyna et al., 2002; Haladyna & Rodriguez, 2013; Muñiz & Fonseca-Pedrero, 2019).

Este primer estudio de validación de la versión en español del TEQ mostró que esta se trata de una escala válida para la evaluación de la empatía en esta población. Además, propone la posibilidad de evaluar el constructo a partir de una escala abreviada de seis ítems. Futuros estudios deberían

confirmar la validez de la escala reducida y analizar si la formulación de ítems invertidos de la escala completa puede estar atentando contra la unidimensionalidad esencial de la misma.

### Agradecimientos

A Juan Pablo Waszuk por la colaboración en la búsqueda y sistematización de la bibliografía.

### Referencias

- Abad, F. J., Olea, J., Ponsoda, V., & García, C. (2011). *Medición en ciencias sociales y de la salud*. Pirámide.
- Arrieta, A., Carballo, M., Bidegain, M., & Ferreira, A. (2021). Hacia un enfoque multinivel en los estudios empíricos de la empatía: Autopercepción, comportamiento y mecanismos neurobiológicos. *Revista de Psiquiatría del Uruguay*, 43-62.
- Auné, S. E., Abal, F. J. P., & Attorresi, H. F. (2017). Propiedades psicométricas de una prueba de conducta empática. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 45(3), 47-56. <https://doi.org/10.21865/RIDEP45.3.04>
- Balbinotti, M. A. A. (2004). Estou testando o que imagino estar? Reflexões acerca da validade dos testes psicológicos. En C. E. Vaz & R. L. Graff (Eds.), *Técnicas Projetivas: Produtividade em Pesquisa* (pp. 6-22, 1.ª Ed.). Casa do Psicólogo.
- Balbinotti, M. A., Benetti, C., & Terra, R. P. (2007). Translation and validation of the Graham-Harvey survey for the Brazilian context. *International Journal of Managerial Finance*, 3(1), 26-48. <https://doi.org/10.1108/17439130710721644>
- Bandalos, D. L., & Finney, S. J. (2018). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R.O. Mueller (Eds.), *The Reviewer's Guide to Quantitative Methods in the Social Sciences*, (pp. 93-114). Routledge. <https://doi.org/10.4324/9781315755649-8>
- Baldner, C., & McGinley, J. J. (2014). Correlational and exploratory factor analyses (EFA) of commonly used empathy questionnaires: New insights. *Motivation and Emotion*, 38, 727-744. <https://doi.org/10.1007/s11031-014-9417-2>
- Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34(2), 163-175. <https://doi.org/10.1023/B:JADD.0000022607.19833.00>
- Batchelder, L., Brosnan, M., & Ashwin, C. (2017). The development and validation of the Empathy Components Questionnaire (ECQ). *PLOS ONE*, 12(1), Article e0169185. <https://doi.org/10.1371/journal.pone.0169185>
- Cliffordson, C. (2001). Parents' judgments and students' self-judgments of empathy: The structure of empathy and agreement of judgments based on the interpersonal reactivity index (IRI). *European Journal of Psychological Assessment*, 17(1), 36. <https://doi.org/10.1027//1015-5759.17.1.36>
- Cliffordson, C. (2002). The hierarchical structure of empathy: Dimensional organization and relations to social functioning. *Scandinavian Journal of Psychology*, 43, 49-59. <https://doi.org/10.1111/1467-9450.00268>
- Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review*, 8(2), 144-153. <https://doi.org/10.1177/1754073914558466>
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *Catalog of Selected Documents in Psychology*, 10, 85.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113-126. <https://doi.org/10.1037/0022-3514.44.1.113>
- Decety, J., & Jackson, P. L. (2004) The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3(2), 71-100. <https://doi.org/10.1177/1534582304267187>
- Decety, J., Bartal, I. B. A., Uzefovsky, F., & Knafo-Noam, A. (2016). Empathy as a driver of prosocial behaviour: Highly conserved neurobehavioural mechanisms across species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1686),

20150077.  
<https://doi.org/10.1098/rstb.2015.0077>
- Elosua-Oliden, P., & Zumbo, B. D. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema*, 20(4), 896-901.
- Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición*, 6(1), 27-36.
- Ferrando, P. J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo*, 31(1), 18-33.
- Ferrando, P. J., & Lorenzo-Seva, U. (2017). Program Factor at 10: Origins, development and future directions. *Psicothema*, 29(2), 236-240.  
<https://doi.org/10.7334/psicothema2016.304>
- Ferrando, P. J., & Lorenzo-Seva, U. (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement*, 78(5), 762-780.  
<https://doi.org/10.1177/0013164417719308>
- Fernández, A. M., Dufey, M., & Kramp, U. (2011). Testing the psychometric properties of the Interpersonal Reactivity Index (IRI) in Chile. *European Journal of Psychological Assessment*, 27(3), 179-185.  
<https://doi.org/10.1027/1015-5759/a000065>
- Flora, D. B., Labrish, C., & Chalmers, R. P. (2012). Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Frontiers in Psychology*, 3, 55.  
<https://doi.org/10.3389/fpsyg.2012.00055>
- Ferguson, C. J. (2016). An effect size primer: A guide for clinicians and researchers. In A. E. Kazdin (Ed.), *Methodological Issues and Strategies in Clinical Research*, (pp. 301-310). American Psychological Association.  
<https://doi.org/10.1037/14805-020>
- Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, 40(7), 3510-3521.  
<https://doi.org/10.1007/s12144-019-00300-2>
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2014). *Multivariate data analysis*. Seventh Edition. Pearson Education Limited.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice ítem-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334.  
[https://doi.org/10.1207/S15324818AME1503\\_5](https://doi.org/10.1207/S15324818AME1503_5)
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.  
<https://doi.org/10.4324/9780203850381>
- Hall, J. A., & Schwartz, R. (2019). Empathy present and future. *The Journal of Social Psychology*, 159(3), 225-243.  
<https://doi.org/10.1080/00224545.2018.1477442>
- Hashimoto, H., & Shiomi, K. (2002). The structure of empathy in Japanese adolescents: Construction and examination of an empathy scale. *Social Behavior and Personality: An International Journal*, 30(6), 593-601.  
<https://doi.org/10.2224/sbp.2002.30.6.593>
- Hernández-Nieto, R. A. (2002). *Contributions to Statistical Analysis*. Universidad de Los Andes.
- Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, 33(3), 307-316.  
<https://doi.org/10.1037/h0027580>
- Hojat, M., Mangione, S., Nasca, T. J., Cohen, M. J. M., Gonnella, J. S., Erdmann, J. B., Veloski, J., & Magee, M. (2001). The Jefferson Scale of Physician Empathy: Development and preliminary psychometric data. *Educational and Psychological Measurement*, 61(2), 349-365.  
<https://doi.org/10.1177/00131640121971158>
- Ilgunaite, G., Giromini, L., & Di Girolamo, M. (2017). Measuring empathy: A literature review of available tools. *Applied Psychology Bulletin*, 280(65), 2-28.
- Kim, H., & Han, S. (2016). A validation study of the Toronto Empathy Questionnaire - Korean Version. *Korean Journal of Clinical Psychology*, 35(4), 809821.  
<https://doi.org/10.15842/kjcp.2016.35.4.009>

- Kline, R. B. (2005). *Methodology in the social sciences. Principles and practice of structural equation modeling (2nd ed.)*. Guilford Press.
- Kourmoussi, N., Amanaki, E., Tzavara, C., Merakou, K., Barbouni, A., & Koutras, V. (2017). The Toronto Empathy Questionnaire: Reliability and validity in a Nationwide sample of Greek teachers. *Social Sciences*, 6(2). <https://doi.org/10.3390/socsci6020062>
- Ledesma, R. D., Ferrando, P. J., & Tosi, J. D. (2019). Uso del Análisis Factorial Exploratorio en RIDEP. Recomendaciones para Autores y Revisores. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 52(3), 173-180. <https://doi.org/10.21865/RIDEP52.3.13>
- Llambí, C., & Piñeiro, L. (2012). *Índice de Nivel Socioeconómico INSE, Revisión anual 2012*. Centro de Investigaciones Económicas (CINVE), Uruguay.
- 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. <https://doi.org/10.6018/analesps.30.3.199361>
- Lloret, S., Ferreres, A., Hernández, A., & Tomás, I. (2017). El análisis factorial exploratorio de los ítems: Análisis guiado según los datos empíricos y el software. *Anales de Psicología*, 33(2), 417-432. <https://doi.org/10.6018/analesps.33.2.270211>
- Lorenzo-Seva, U., & Ferrando, P.J. (2006) FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods*, 38, 88-91. <https://doi.org/10.3758/BF03192753>
- Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. (2011). The Hull Method for Selecting the Number of Common Factors. *Multivariate Behavioral Research*, 46(2), 340-364. <https://doi.org/10.1080/00273171.2011.564527>
- Manavipour, D., & Memarian, E. (2020). Toronto Empathy Questionnaire: Evaluation of psychometric properties in a sample of Iranian students. *Archives of Pharmacy Practice*, 1, 106.
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519. <https://doi.org/10.2307/2334770>
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103(3), 391-410. <https://doi.org/10.1037/0033-2909.103.3.391>
- Mehrabian, A. (2000). Beyond IQ: Broad-based measurement of individual success potential or "emotional intelligence." *Genetic, Social, and General Psychology Monographs*, 126, 133-239.
- Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality*, 40(4), 525-543. <https://doi.org/10.1111/j.1467-6494.1972.tb00078.x>
- Montero, I., & León, O. G. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology*, 7(3), 847-862.
- Muñiz, J. (2010). Las Teorías de los Tests: Teoría clásica y teoría de respuesta a los ítems. *Papeles del Psicólogo*, 31(1), 57-66.
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test [Diez pasos para la construcción de un test]. *Psicothema*, 31(1), 7-16. <https://doi.org/10.7334/psicothema2018.291>
- Novak, L., Malinakova, K., Mikoska, P., van Dijk, J. P., Dechterenko, F., Ptacek, R., & Tavel, P. (2021). Psychometric analysis of the Czech version of the Toronto Empathy Questionnaire. *International Journal of Environmental Research and Public Health*, 18(10), 5343. <https://doi.org/10.3390/ijerph18105343>
- Nunnally, J. C. (1978) *Psychometric theory*. 2nd Edition, McGraw-Hill
- Ondé, D., & Alvarado, J. M. (2022). Contribución de los modelos factoriales confirmatorios a la evaluación de estructura interna desde la perspectiva de la validez. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 66(5), 5. <https://doi.org/10.21865/ridep66.5.01>
- Pérez, C., & Carretero-Dios, H. (2005). Normas para el desarrollo y revisión de estudios instrumentales. *International Journal of Clinical and Health Psychology*, 5(3), 521-551.

- Qualtrics (2021) Provo, UT, USA.  
<https://www.qualtrics.com/blog/citing-qualtrics>
- Reynolds, W., Scott, P. A., & Austin, W. (2000). Nursing, empathy and perception of the moral. *Journal of Advanced Nursing*, 32, 235-242.  
<http://dx.doi.org/10.1046/j.1365-2648.2000.01440.x>
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323-338.  
<https://doi.org/10.3200/JOER.99.6.323-338>
- Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences*, 25, 167-177.  
[https://doi.org/10.1016/S0191-8869\(98\)00001-4](https://doi.org/10.1016/S0191-8869(98)00001-4)
- Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment*, 91(1), 62-71.  
<https://doi.org/10.1080/00223890802484381>
- Suárez-Álvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed ítems in Likert scales: A questionable practice. *Psicothema*, 30(2), 149-158.  
<https://doi.org/10.7334/psicothema2018.33>
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous ítems with parallel analysis. *Psychological Methods*, 16(2), 209-220.  
<https://doi.org/10.1037/a0023353>
- Totan, T., Doğan, T., & Sapmaz, F. (2012). The Toronto empathy questionnaire: Evaluation of psychometric properties among Turkish university students. *Egitim Arastirmalari - Eurasian Journal of Educational Research*, 46, 179-198.
- Ventura-León, J. L., & Caycho-Rodríguez, T. (2017). El coeficiente Omega: Un método alternativo para la estimación de la confiabilidad. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 15(1), 625-627.
- Wang, Y.-W., Davidson, M. M., Yakushko, O. F., Savoy, H. B., Tan, J. A., & Bleier, J. K. (2003). The Scale of Ethnocultural Empathy: Development, validation, and reliability. *Journal of Counseling Psychology*, 50(2), 221-234.  
<https://doi.org/10.1037/0022-0167.50.2.221>
- Woods, C. M. (2006). Careless responding to reverse-worded ítems: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28(3), 189-194.  
<https://doi.org/10.1007/s10862-005-9004-7>
- Yeo, S., & Kim, K. J. (2021). A validation study of the Korean version of the Toronto empathy questionnaire for the measurement of medical students' empathy. *BMC Medical Education*, 21(1).  
<https://doi.org/10.1186/s12909-021-02561-7>
- Xu, R. H., Wong, E. L. yi, Lu, S. Y. jun, Zhou, L. M., Chang, J. H., & Wang, D. (2020). Validation of the Toronto Empathy Questionnaire (TEQ) among medical students in China: Analyses using three psychometric methods. *Frontiers in Psychology*, 11, 1-11.  
<https://doi.org/10.3389/fpsyg.2020.00810>