

Análisis psicométrico del Cuestionario de Empatía de Toronto, aplicado a una muestra española

Juan Manuel RODRÍGUEZ JIMÉNEZ

Gabinete de Psicología Aplicada y Análisis de Conducta, Gijón, Asturias (España)

María Teresa VEGA RODRÍGUEZ

Universidad de Salamanca (España)

Resumen

La medición de la empatía presenta problemas asociados tanto a la propia definición, como a la naturaleza del constructo, así como a la escasez de investigación sobre el mismo. Como consecuencia, los profesionales disponen de un número limitado de pruebas psicométricas dedicadas a su medición. Por ello, el presente trabajo tiene como finalidad analizar las propiedades psicométricas de una versión en español del *Toronto Empathy Questionnaire*, un instrumento diseñado para medir de forma unidimensional la empatía, así como su adecuación para su uso en población española en actividades de investigación. Ha sido aplicado a una muestra heterogénea compuesta por 573 sujetos, 65% mujeres ($N=373$) y 35% hombres ($N=201$). Los resultados indicaron que para obtener una prueba parsimoniosa se debería crear una escala reducida de 11 ítems, en lugar de los 16 ítems de la escala original. La nueva escala no sería estrictamente unidimensional, debido a que en el Análisis Factorial Exploratorio fueron aislados cuatro componentes, agrupados en torno a tres dimensiones diferentes.

Palabras clave: empatía, Cuestionario de Empatía de Toronto, indiferencia emocional, sintonía emocional.

Abstract

The measurement of empathy presents problems associated with both the definition itself and the nature of the construct, as well as the scarcity of research on it. As a consequence, professionals have a limited number of psychometric tests dedicated to its measurement. The purpose of this paper is to analyze the psychometric properties of the Toronto Empathy Questionnaire, a tool for measuring empathy in a unidimensional way, as well as its suitability for use in the Spanish population in research activities. It has been applied to a heterogeneous sample made up of 573 subjects, 65% women ($N=373$) and 35% men ($N=201$). The results indicated that to obtain a parsimonious test, a reduced scale of 11 items should be created, instead of the 16 items of the original scale. The new scale would not be strictly unidimensional, because in the Exploratory Factor Analysis four components were isolated, grouped around three different dimensions.

Key words: empathy, TEQ, Toronto Empathy Questionnaire, emotional indifference, emotional connection

La empatía conforma un constructo en el que diferentes dimensiones y componentes entran en escena, dificultando el consenso entre los estudiosos del tema y la creación de un cuerpo teórico sólido que facilite el desarrollo de pruebas psicométricas adecuadas.

Las diferentes propuestas se han basado en procesos emocionales y sentimientos subjetivos, tales como el sentimiento de felicidad como favorecedor de la empatía (Seligman, 2004), características comunes entre perso-

nas como factores facilitadores de la conducta empática (Batson, Turk, Shaw y Klein, 1995) o procesos cognitivos facilitadores de una comprensión racional (Lawrence, Shaw, Baker *et al.*, 2004) en oposición a procesos emocionales independientes de los cognitivos (Rankin, Kramer y Miller, 2005).

Otros autores (Fernández-Pinto, López-Pérez y Márquez, 2008) proponen modelos en los que se dan procesos cognitivos y emocionales de forma conjunta, clasificán-

dolos en diferentes grupos: empatía cognitivo-afectiva de signo negativo, empatía cognitivo-emocional de signo positivo, contagio emocional y empatía cognitiva, a la vez que se también se incide en las variables disposicionales, ya que, de otra forma, el modelo quedaría incompleto.

La investigación reciente pone énfasis en los componentes cognitivos y afectivos, asumiendo varias definiciones, por lo que el acuerdo entre investigadores es difícil de alcanzar (Spreng, McKinnon, Mar y Levine, 2009). Como consecuencia de ello, tampoco existe consenso en el modo en que se debe evaluar la empatía, considerando los diferentes enfoques existentes y, por ende, constructos subyacentes en los cuestionarios.

De este modo, existen pruebas basadas en una visión integradora, como el Test de Empatía Cognitiva y Afectiva (TECA), de López-Pérez, Fernández-Pinto y García (2008) o el Índice de Reactividad Interpersonal (IRI), desarrollado por Davis (1983). Desde una perspectiva afectiva, se desarrolló la Escala de Empatía Emocional Balanceada (Mehrabian, 1997), en base a variables cognitivas, como la Escala de Empatía (Hogan, 1969), o incluso escalas situacionales, como la Escala de Empatía e Identificación con los personajes (Igartua Perosanz y Paez Rovira, 1998).

A pesar de la existencia de múltiples desarrollos, y aunque en algunas situaciones sea preferible una medida multifacética, la medición de este constructo carece de una herramienta adecuada en un nivel amplio (Spreng *et al.*, 2009). Con esta finalidad se creó el Cuestionario de Empatía de Toronto (*Toronto Empathy Questionnaire*, TEQ), desarrollado por Nathan Spreng y colaboradores (2009). La herramienta pretende ser un complemento de las pruebas psicométricas multifactoriales de uso habitual. Su desarrollo se produjo a través de tres estudios diferentes. En el primero, se llevó a cabo un análisis factorial exploratorio en base a la recopilación de autoinformes sobre todas las medidas de empatía posibles. Los participantes fueron 200 estudiantes de la Universidad de Toronto, estando distribuida la muestra al cincuenta por ciento entre hombres y mujeres. Las preguntas evaluaron la frecuencia del comportamiento en lugar de tendencias, para lo que se usó una escala Likert correspondiente a cinco niveles de frecuencia, donde 0 es “nunca” y 4 “siempre”.

En un segundo estudio, participaron 79 estudiantes de la misma universidad. A través de la prueba Test de la Mirada (MIE) (Baron-Cohen, Wheelwright, Hill *et al.*, 2001), empleada para detectar la capacidad de comprensión sobre estados mentales por medio de señales no verbales, la *Interpersonal Perception Task* (IPT-15) (Archer y Costanzo, 1993), en la que a través de la observación de interacciones en vídeo la persona debe atender a las señales no verbales para dar una respuesta correcta a preguntas de opción múltiple, y el cuestionario IRI (Davis, 1980), se correlacionaron las puntuaciones totales para examinar la validez del TEQ.

El tercer estudio, tuvo como finalidad profundizar en las propiedades psicométricas del TEQ a través de un nuevo análisis de validez convergente y discriminante. Para ello, fueron empleadas medidas de autoinforme sobre empatía y sintomatología del espectro autista, junto con datos obtenidos sobre fiabilidad test-retest en base a los participantes del segundo estudio, añadiendo nuevas medidas de procesamiento social y el Cociente de Empatía (Baron-Cohen y Wheelwright, 2004).

Los resultados mostraron correlaciones positivas entre el TEQ y las escalas de preocupación empática y toma de perspectiva del IRI y con las medidas de comprensión social del IPT-5 y el MIE. Las medidas del MIE y escalas del IRI no mantienen correlación, o esta es negativa, al igual que ocurre con las medidas del MIE y el IPT-5. Esta falta de correlación entre medidas que pretenden abordar un mismo constructo es muestra de la heterogeneidad que caracteriza la medición de la empatía (Spreng *et al.*, 2009). Precisamente, los autores proponen el TEQ como una herramienta que puede suplir la carencia de una medida unificada de la empatía, poniendo énfasis en los componentes emocionales de tipo cognitivo y afectivo.

El cuestionario consta de 16 ítems, en los que los autores indican una asociación con elementos teóricos, tales como la comprensión de las emociones, el contagio emocional, altruismo y aspectos asociados a la excitación fisiológica simpática. Así mismo, se demuestra una asociación con la explicación cognitiva de la teoría de la mente, al encontrarse correlaciones con las subescalas del IRI “toma de perspectiva” y “fantasía” que se describen como componentes de carácter cognitivo en la empatía (Davis, 1983).

La aplicación del cuestionario sobre la muestra seleccionada adolece de limitaciones respecto al rango de edad, al ser administrada a población universitaria. No obstante, los autores sugieren que el valor intrínseco de la prueba se encuentra precisamente en constituirse como una propuesta adecuada con la que aislar y hacer manejable un constructo que, por su propia naturaleza, responde a una fenomenología compleja y multidimensional.

Tal como los mismos autores reflejan en su investigación, no es intención establecer una medida única de la empatía, sino que el objetivo es desarrollar una herramienta poco extensa y parsimoniosa que complemente otras medidas de empatía de modo que facilite su comprensión en contextos, tanto clínicos como de investigación, haciendo manejable su aplicación.

El cuestionario ha sido validado en diferentes poblaciones en muestras pertenecientes a colectivos concretos. En Corea (Yeo y Kim, 2021), el estudio se efectuó sobre una muestra de 279 estudiantes de medicina, a través de análisis factorial confirmatorio y exploratorio, así como probando la bondad de ajuste, obteniéndose un alfa de

entre 0'71 y 0'81. Dos ítems fueron eliminados, quedando la prueba reducida a 14 ítems.

El estudio llevado a cabo sobre población china (Xu, Wong, Lu *et al.*, 2020), consistió en una muestra de 1296 estudiantes de medicina, aplicándose técnicas basadas en la Teoría Clásica de los Test, Teoría de Respuesta al Ítem y el Modelo de Rasch. Para valorar la fiabilidad se empleó la técnica test-retest, sobre una muestra de 40 estudiantes en fase retest, realizada dos semanas después de la primera aplicación. Los resultados encontraron a través de análisis factorial tres factores, indicando los investigadores la necesidad de que algunos elementos deberían ser modificados y revisados.

La validación sobre población griega (Kourmousi, Amanaki, Tzavara *et al.*, 2017) se realizó sobre una muestra compuesta por 3955 profesores. Se efectuó un análisis factorial exploratorio sobre la muestra par y confirmatorio sobre la impar. El modelo presentó un buen ajuste, señalando diferencias en variables tales como el sexo, haber recibido cursos de capacitación, estado laboral, etc.

En Turquía (Totan, Doğan y Sapmaz, 2012) se validó sobre una muestra de 698 estudiantes a través del análisis de ítems, análisis factorial confirmatorio y exploratorio, validez de criterio y aplicación test-retest. El estudio concluyó que se mantuvo la unidimensionalidad de la muestra, encontrándose así mismo correlaciones positivas con las pruebas Escala de Tendencia Empática y la Escala Básica de Empatía. El coeficiente de fiabilidad test-retest fue de 0'73 y el de consistencia interna fue 0'79.

Otros estudios de validación fueron desarrollados en países latinoamericanos. En Perú (Carranza Albán y Huamani Urbina, 2020), se realizó sobre una muestra de 414 universitarios, obteniéndose dos factores, resonancia empática y distancia empática, con un alfa de Cronbach de 0'80 y 0'79, respectivamente. Así mismo, presentó una adecuada validez en la estructura interna.

Un nuevo estudio (Muñoz y Valenzuela, 2016) fue realizado en base a la validación previa realizada en adolescentes chilenos. La muestra estuvo compuesta por 386 sujetos, divididos en dos grupos. El primero compuesto por 340 adolescentes sin conducta transgresora o antisocial y el segundo grupo compuesto por 46 adolescentes, integrados en programas del Servicio Nacional de Menores (SENA-ME), perteneciente al Ministerio de Desarrollo Social y Familia del Gobierno de Chile. Los autores obtuvieron una fiabilidad aceptable (alfa de Cronbach 0'80). La validez fue determinada de forma concurrente con el TECA (López-Pérez, Fernández-Pinto y Abad García, 2008), obteniendo $r = 0'68$ ($p = 0'00$). Los autores indicaron que el Cuestionario de Empatía de Toronto era una herramienta adecuada para su aplicación de cribado en adolescentes chilenos.

El presente estudio tiene un doble objetivo. En primer lugar, analizar las propiedades psicométricas de la prueba,

traducida, adaptada y aplicada a una muestra de la población española con la finalidad de conocer su adecuación. El segundo objetivo es llevar a cabo una selección de ítems fiables para el desarrollo de nuestra investigación en el campo de la empatía y su estimulación en diferentes contextos psicosociales.

En este sentido, los resultados obtenidos ponen de relieve la dificultad que conlleva el desarrollo de una prueba suficientemente parsimoniosa como para captar de un modo unidimensional la naturaleza de un constructo complejo como es la empatía. Esto se debe a que existe una dificultad que entraña el ajuste a los modelos teóricos establecidos debido a la dispersión y disparidad de enfoques por el propio diseño y desarrollo de las pruebas o por la naturaleza multi-componente del constructo. Todo ello dificulta el desarrollo de un corpus teórico unificado.

Es habitual que la investigación se encuentre con áreas que se solapan unas a otras, falta de fiabilidad o dificultades en la operativización debido a la mencionada dispersión de componentes. Precisamente, es lo que la prueba TEQ pretende solucionar, razón por la que fue elegida como prueba de referencia en la presente investigación, en conjunción con el TECA, ya adaptado a población española.

Método

Muestra

La muestra estuvo formada, una vez eliminados los sujetos que contestaron al cuestionario y no cumplían los criterios de edad o no proporcionaron datos causando mortalidad experimental, por 574 personas, procedentes de diversas provincias de España y diferentes ocupaciones laborales.

La distribución por sexos fue de un 65% mujeres ($N=373$) y un 35% hombres ($N=201$), presentando la edad una ligera distribución asimétrica negativa (-0,37) bimodal leptocúrtica (-0,685). La media de edad es de 42 años (41,70). La edad mínima es de 18 años y la máxima de 77.

Instrumentos de medida

Con el objetivo de medir la empatía, entendida como el conjunto de procesos cognitivos y afectivos que capacitan a una persona para ponerse en el lugar de otra, se aplicó el cuestionario TEQ (tabla 1) que consta de 16 ítems. Como puede apreciarse en su redacción, los ítems, 1, 3, 5, 6, 8, 9, 13, 16 son directos y, en consecuencia, se califican de modo creciente (Nunca = 0; Rara vez = 1; A veces = 2; A menudo = 3 Siempre = 4), en tanto que los ítems inversos 2, 4, 7, 10, 11, 12, 14, 15 se puntúan de modo decreciente (Nunca = 4; Rara vez = 3; A veces = 2; A menudo = 1; Siempre = 0).

Tabla 1. Cuestionario de Empatía de Toronto. Versión adaptada al español del TEQ (Spreng *et al.*, 2009).

1. Cuando alguien se siente emocionado, yo también suelo emocionarme con él/ella, es como si sintiera lo mismo (por ejemplo, si alguien se pone triste, me pongo triste; si está contento/a, me pongo contento/a).
2. Las desgracias de otros/as no me molestan mucho, me dan igual.
3. Me molesta ver que alguien está siendo tratado/a de manera irrespetuosa.
4. Cuando alguien cercano a mí está feliz, me da igual, no siento nada en especial.
5. Disfruto haciendo que otras personas se sientan mejor.
6. Siento afecto y me preocupo por las personas menos afortunadas que yo.
7. Cuando un amigo o amiga empieza a hablar de sus problemas, trato de dirigir la conversación hacia otra cosa (cambiar de tema).
8. Puedo saber cuándo los demás están tristes incluso cuando no dicen nada.
9. Pienso que estoy “en sintonía” y con los estados de ánimo de otras personas, es decir, puedo sentirme como ellos/as se sienten.
10. No siento simpatía por las personas que causan sus propias enfermedades graves (que tienen enfermedades por su culpa).
11. Me enfado cuando alguien llora.
12. No estoy interesado/a en cómo se sienten los demás (me da igual cómo se sienten).
13. Siento un fuerte impulso de ayudar a una persona cuando veo que se encuentra mal o está molesta.
14. Cuando alguien está siendo tratado/a injustamente, me da igual y no siento lástima por él/ella.
15. Me parece una tontería que la gente lllore por felicidad.
16. Me dan ganas de ayudar y proteger a una persona cuando veo que se aprovechan o quieren aprovecharse de él/ella.

Procedimiento

En primer lugar, se contactó con el autor de la prueba, Nathan Spreng, con la finalidad de obtener su autorización para usar y reproducir la prueba en condiciones experimentales e investigación, de acuerdo con las directrices de la Comisión Internacional de Tests (Hambleton y Li, 2005; Muñiz, Elosua y Hambleton, 2013). A continuación, se procedió a la traducción al idioma español, participando para ello dos psicólogas con dominio del idioma inglés y con experiencia idiomática, al haber sido residentes en países de lengua inglesa. En la traducción de los ítems se tuvo especial cuidado en mantener lo máximo posible la literalidad original, pero algunos ítems tuvieron que ser ligeramente modificados debido a que podrían plantear dudas o dificultad en la interpretación de los mismos. Las modificaciones consistieron en añadir notas aclaratorias, entre paréntesis, sobre el sentido de las cuestiones planteadas.

Debido a la situación de pandemia y la imposibilidad de llevar a cabo una administración presencial, se procedió a crear una aplicación online para su distribución. La captación se llevó a cabo utilizando redes sociales, solicitando colaboración en cadena. La prueba fue camuflada como “sondeo de opinión” sobre una investigación con el objetivo de minimizar sesgos en las respuestas en relación a posibles efectos sobre la deseabilidad social a la vez que se solicitó a los participantes la mayor sinceridad posible. En todo momento, se garantizó el anonimato al no extraerse ningún dato identificativo personal, salvo datos socio-demográficos de interés para la investigación tales como sexo, edad, residencia y actividad profesional, ninguno de los cuales permiten una identificación sobre la persona que responde al cuestionario.

El cuestionario se mantuvo en línea durante un período de 22 días, del 15 de noviembre al 6 de diciembre de 2020. Los datos fueron recogidos directamente en la hoja de cálculo de la aplicación y procesados con IBM SPSS Statistics versión 25.

Resultados

Análisis de los datos

Respecto al análisis efectuado sobre el cuestionario original (16 ítems), se ha obtenido un alfa de Cronbach de 0,801 (en elementos estandarizados, $\alpha = 0,81$). Por medio de la técnica de las dos mitades, los coeficientes de Spearman-Brown y Guttman son de 0,81. La correlación entre formularios es de 0,70 y el coeficiente alfa es de 0,64 para la primera parte y 0,67 para la segunda. El alfa de Cronbach en el TEQ original, reportado por los autores a través del primer estudio fue de 0,85, por lo que el coeficiente obtenido en nuestra muestra, aunque más bajo ($\alpha = 0,80$), es aproximado a la administración original.

Con el objetivo de poder valorar la unidimensionalidad teórica asumida en el TEQ, se procedió a realizar la reducción del conjunto de variables a factores, establecidos en base a la percepción de los participantes sobre los ítems del cuestionario. El método de extracción es a través de componentes principales, optándose por rotación ortogonal varimax con normalización Kaiser, suprimiendo aquellos coeficientes cuyo valor absoluto sea inferior a 0,40.

El índice KMO en la prueba con la muestra total es satisfactorio, 0,849 y Chi-cuadrado $\chi^2 = 2320,94$ ($gl = 120$; $p = 0,00$). En la matriz anti-imagen de la covarianza, que contiene los negativos de las covarianzas parciales, los elementos

no diagonales son bajos. La matriz de las correlaciones anti-imagen, que contiene los negativos de los coeficientes de correlación parcial, tiene valores pequeños también fuera de las diagonales. Por otro lado, en la matriz anti-imagen se puede observar que la Medida de Adecuación a la Muestra (MSA) contiene en la diagonal valores altos. La prueba de esfericidad de Barlett nos lleva a rechazar la hipótesis de que la matriz de correlaciones es la matriz identidad, por lo que la muestra es adecuada para el análisis factorial.

El AFE arroja cuatro componentes que explican el 54'5% de la varianza común. Los dos primeros componentes explican el 40'49% de la varianza. El tercer componente explica el 7'61% y el cuarto el 6'40%.

Ítems formulados mediante negaciones en la versión original

El cuestionario contiene cinco ítems formulados a través de negaciones (ítems 2, 4, 10, 12 y 14). El AFE efectuado sobre estos cinco ítems arroja un solo componente, que aunque responde a un factor común cuya naturaleza es la "indiferencia ante sentimientos de los otros", son eliminados en nuestro modelo debido a que la Medida de Adecuación a la Muestra no es satisfactoria, aun siendo significativa la prueba de esfericidad de Bartlett, con $\chi^2 = 397'37$ ($p = 0'00$), KMO es de 0'74 y la configuración de la matriz de correlaciones anti-imagen arroja coeficientes moderados en la diagonal. Por otro lado, el alfa de Cronbach para estos cuatro ítems en la muestra es de 0'64.

Eliminando los ítems formulados en negativo del modelo, el cuestionario queda conformado por 11 ítems. El índice KMO es de 0'83 arrojando la prueba de esfericidad

de Barlett, $\chi^2 = 1614'20$ ($p = 0'00$), obteniendo tres componentes (véase la tabla 2).

El porcentaje acumulado de varianza explicada del 57'08% correspondiendo un 35'12% al primer componente, 12'78% al segundo y 9'18% al tercero. El alfa de Cronbach para estos 11 elementos es de 0'797.

Estos tres componentes se encuentran relacionados con los siguientes patrones: el componente 1 está compuesto por los ítems 3, 5, 6, 13 y 16. Estos ítems están asociados a conductas prosociales, dirigidas a la protección, defensa y/o ayuda de personas que se encuentran en situaciones desfavorecidas, bien porque están siendo molestadas o maltratadas. Por su parte, el componente 2, estaría formado por los ítems 1, 6, 8 y 9. Dichos ítems están relacionados con la capacidad para reconocer emociones y estados de ánimo en los otros. En cuanto al componente 3, lo conforman los ítems 7, 11 y 15, que estarían relacionados con la falta de interés en compartir emociones.

La consistencia interna para cada uno de los componentes por separado es de $\alpha = 0'799$ para el primero, $\alpha = 0'70$ para el segundo y $\alpha = 0'532$ para el tercer componente. Respecto a este último componente o factor, la consistencia interna es poco aceptable, pero cualitativamente está vinculado a un factor diferenciado de los demás.

Análisis de ítems directos e inversos

En su versión original el cuestionario contiene la mitad de los ítems medidos de forma inversa (8 ítems: 2, 4, 7, 10, 11, 12, 14 y 15). Como resultado del análisis realizado sobre este grupo, se obtiene KMO 0'82, siendo $\chi^2 = 694'22$ ($p = 0'000$). La consistencia interna es $\alpha = 0'702$ ($\alpha =$

Tabla 2. Análisis factorial de la escala de 11 ítems (versión TEQ 11). Matriz de componente rotado.

	Componente		
	1	2	3
1. Cuando alguien se siente emocionado, yo también suelo emocionarme con él/ella, es como si sintiera lo mismo (por ejemplo, si alguien se pone triste, me pongo triste; si está contento/a, me pongo contento/a).		.744	
3. Me molesta ver que alguien está siendo tratado/a de manera irrespetuosa.	.699		
5. Disfruto haciendo que otras personas se sientan mejor.	.684		
6. Siento afecto y me preocupo por las personas menos afortunadas que yo.	.611	.410	
7. Cuando un amigo o amiga empieza a hablar de sus problemas, trato de dirigir la conversación hacia otra cosa (cambiar de tema).			.649
8. Puedo saber cuándo los demás están tristes incluso cuando no dicen nada.		.595	
9. Pienso que estoy "en sintonía" y con los estados de ánimo de otras personas, es decir, puedo sentirme como ellos/as se sienten.		.836	
11. Me enfado cuando alguien llora.			.740
13. Siento un fuerte impulso de ayudar a una persona cuando veo que se encuentra mal o está molesta.	.719		
15. Me parece una tontería que la gente lllore por felicidad.			.696
16. Me dan ganas de ayudar y proteger a una persona cuando veo que se aprovechan o quieren aprovecharse de él/ella.	.792		

0'715 en elementos estandarizados). El análisis factorial aísla dos factores, constituidos por cinco ítems en el factor 1 y cinco en el factor 2, solapándose los ítems 12 y 14 en ambos (el número de ítems que constituye la escala es de ocho). El factor 1 está vinculado a la falta de interés sobre los “estados emocionales de los otros”, conteniendo dos ítems formulados con negación. El factor 2, se encuentra relacionado con posiciones sobre “indiferencia ante los estados emocionales”, constituyendo este factor todos los ítems de la escala formulados con negación.

Respecto a los ítems medidos de forma directa (ocho ítems: 1, 3, 5, 6, 8, 9, 13 y 16), en el análisis factorial realizado se obtiene KMO 0'828 y $\chi^2 = 1386'37$ ($p = 0'000$), obteniendo dos factores constituidos por cinco ítems cada uno, que contienen dos ítems en común (ítems 6 y 13). El primer factor ($\alpha = 0'779$) está vinculado a las “conductas de ayuda y preocupación por los demás”, mientras que el segundo ($\alpha = 0'759$) contiene de un modo directo una carga más emocional, encontrándose relacionado con la “conexión y sintonía en relación a las emociones de los otros”. Es decir, compartir y sentir esas emociones. La consistencia interna de la escala conformada por los ocho ítems es $\alpha = 0'82$ ($\alpha = 0'82$ en elementos estandarizados).

Eliminados los ítems comunes a los dos factores mencionados en la escala conformada por los ítems directos, se obtiene una escala de seis ítems con dos factores (tres ítems cada uno) completamente diferenciados, con un porcentaje de varianza explicada del 62'66%, (45'06% primer factor y 17,60% el segundo) con $\alpha = 0'746$. Se mantiene la estructura teórica anteriormente comentada, factor 1 ($\alpha = 0'717$), “conductas de ayuda y preocupación por los demás”, y factor 2, ($\alpha = 0'663$) “sentir las emociones de los otros”.

Con el objetivo de conseguir un solo factor, se optó por eliminar un ítem, con la condición de que la escala saturase en un solo factor que explique el mayor porcentaje de varianza junto con la mayor consistencia interna posible. De este modo, eliminando uno de ellos y manteniendo los restantes, con lo que mantenemos una escala de cinco ítems, los datos obtenidos fueron: eliminando el ítem 1, se arroja un coeficiente alfa de Cronbach $\alpha = 0'74$ y un factor, siendo la varianza explicada por éste un 49'82%. Eliminando el

ítem ítem 3, la escala queda con $\alpha = 0'70$ y un solo factor, que explica el 46,09% de varianza; eliminando el ítem 9, $\alpha = 0'69$ y un solo factor, con 46'64% de varianza explicada; para el ítem 16, $\alpha = 0'71$, un solo factor y un 47'23% de varianza. En cambio, si eliminamos en nuestra escala de 6 ítems el ítem 5, con $\alpha = 0'707$ se producen dos factores, explicando el 46'74% de varianza el primero y un 20'68% el segundo. El porcentaje acumulado es del 66'80%.

También se originan dos factores al eliminar el ítem 8, obteniéndose un coeficiente $\alpha = 0'71$. El primer factor explica una varianza del 47'51% y un 20'69% el segundo, siendo 68'20% el valor acumulado. Se obtienen dos factores en cada una seleccionando aquél que maximizase la consistencia interna de la nueva escala. En el empeño para la reducción de factores, el ítem seleccionado para ser eliminado fue el número 1, decisión basada en que la configuración que resulta de su eliminación es la que muestra saturación en un solo factor con una mayor consistencia interna ($\alpha = 0'74$) junto con un mayor porcentaje de varianza explicada por ese factor (49'82%). Por tanto, la escala finalmente quedó conformada por los ítems 3, 5, 8, 9 y 16 (tabla 3).

Diferencias por sexo

La puntuación directa total que ofrece la prueba oscila entre 0 y 64. La media obtenida en el conjunto de la población es de 50'82. La prueba t para una muestra da como resultado $t = -0'017$ ($gl = 573$; $p = 0'99$), por lo que podemos concluir que la media se aproximaría a la poblacional. La correspondiente a mujeres es de 52'27 y la de hombres 48'11.

Con el fin de comprobar si estas medias son estadísticamente significativas, se sometieron a una prueba t para medias independientes, con un 99% de intervalo de confianza para las diferencias, obteniendo los siguientes resultados: $F = 7'268$ ($p = 0'007$; $t = -7'324$ ($gl = 572$; $p = 0'00$), por lo que se asumen varianzas diferentes para ambos grupos. Siendo en este caso (diferentes varianzas) $t = -6'845$ ($gl = 572$; $p = 0'00$), se puede concluir que las medias respectivas de ambos grupos son diferentes, obteniendo las mujeres una mayor puntuación global.

Tabla 3. Escala unifactorial, versión de 5 ítems. Método de extracción: análisis de componentes principales. Matriz del componente.

	<i>Componente 1</i>
3. Me molesta ver que alguien está siendo tratado/a de manera irrespetuosa.	.750
5. Disfruto haciendo que otras personas se sientan mejor.	.716
8. Puedo saber cuándo los demás están tristes incluso cuando no dicen nada.	.682
9. Pienso que estoy “en sintonía” y con los estados de ánimo de otras personas, es decir, puedo sentirme como ellos/as se sienten.	.659
16. Me dan ganas de ayudar y proteger a una persona cuando veo que se aprovechan o quieren aprovecharse de él/ella.	.719

Igualmente, las mujeres obtienen mayor puntuación media en todos los ítems, valorados uno a uno. Los ítems con varianzas diferentes son: 1, 3, 5, 8, 15 y 16, siendo 2, 4, 6, 7, 9, 10, 11, 12, 13 y 14 los ítems con varianzas iguales. En todos los ítems existen diferencias significativas entre las medias ($p = 0'001$), salvo en los ítems 2, 4 y 10, en los que las diferencias no son significativas ($p = 0'18$, $p = 0'14$ y $p = 0'44$, respectivamente), conformando tres de los cinco ítems formulados a través de negaciones (en el ítem número 12, sí se aprecian diferencias estadísticamente significativas entre hombres y mujeres).

Centrado el análisis sobre los ítems directos, la media en el total de puntuación es de 24'69, siendo 23'09 en hombres y 25'56 en mujeres. La media total resulta significativa a nivel poblacional con $t = -0'001$ ($gl = 573$; $p = 0'05$).

Discusión

A través de los datos obtenidos sobre fiabilidad en nuestra muestra, y aplicando los 16 ítems, se obtiene un valor de $\alpha = 0'80$, relativamente próximo al obtenido en la validación original ($\alpha = 0'85$) y a los estudios realizados para validar la prueba en poblaciones específicas de Corea (entre $\alpha = 0'71$ y $\alpha = 0'81$), Turquía ($\alpha = 0'79$), Perú ($\alpha = 0'80$ y $\alpha = 0'79$), o Chile ($\alpha = 0'80$). Por tanto, se puede asumir que la prueba dispone de un nivel aceptable en consistencia interna y fiabilidad. No obstante, se observa que los índices de homogeneidad de los ítems son bajos, oscilando entre 0'18 y 0'57. Así mismo, se entiende que un test es unidimensional si es fiable, es decir, si mide adecuadamente el constructo que se pretende medir en base a la varianza compartida, intercorrelación entre ítems y la homogeneidad entre éstos, es decir, la similitud entre dichas intercorrelaciones.

Otros autores refieren el término homogeneidad como directamente equivalente a unidimensionalidad. En la muestra del presente estudio, los índices de homogeneidad, son bajos o moderados. El uso de ítems formulados a través de preguntas con negaciones suele plantear problemas de interpretación en las personas que ejecutan las pruebas psicométricas. En el primer análisis realizado, y con el fin de mantener la integridad de la misma, se mantuvieron tal y como estaban en la prueba original. En un segundo análisis, y con el fin de valorar con la mayor exactitud posible sus propiedades, se procedió a eliminar los ítems formulados con negación (cinco ítems), obteniendo una escala de 11 elementos. Por esta razón, el coeficiente alfa descendió ligeramente, a 0'795, observando que la correlación total entre elementos corregida se sitúa entre un mínimo de 0'19 y un máximo de 0'62. La reducción de factores arroja tres componentes, compuestos por cinco ítems el primero, cuatro el segundo y tres el tercero.

La correlación total entre elementos corregida se sitúa entre 0'55 el menor valor y 0'66 el mayor. Respecto al segundo componente, los índices de homogeneidad bajan

a 0'44 el menor y 0'59 el mayor. Respecto al tercero, se produce una menor calidad en los índices, situándose entre 0'332 el menor y 0'37 el mayor. El coeficiente alfa es aceptable para la escala del primer componente ($\alpha = 0'80$), e incluso para el segundo componente ($\alpha = 0'70$), pero baja para el tercero ($\alpha = 0'53$).

En relación al análisis cualitativo de estos tres componentes, se aprecia claramente que los ítems que conforman cada una de estas subescalas se agrupan en ellas en torno a determinados aspectos comunes, existiendo diferencias cualitativas entre ellas.

De este modo, los componentes vienen definidos por las siguientes características:

- Componente 1: vinculado a molestia percibida cuando el otro se siente mal.
- Componente 2: asociado al reconocimiento y sintonía con las emociones.
- Componente 3: indiferencia ante las emociones de los otros.

La prueba original ha sido desarrollada conformando una escala en la que la mitad de los ítems puntúan de 0 a 4 el grado de identificación que siente la persona con las situaciones presentadas (0: nunca, 1: raramente, 2: a veces, 3: a menudo y 4: siempre) y la otra mitad son ítems que puntúan de modo inverso (0: siempre, 1: menudo, 2: a veces, 3: raramente y 4: nunca).

Los ítems que puntúan de modo inverso no superan el coeficiente de .8, situándose en un $\alpha = 0'70$. Aunque la diferencia con la escala compuesta por ítems directos ($\alpha = 0'82$) no es excesivamente amplia, sí es lo suficientemente grande como para plantear dudas sobre la conveniencia de introducir ítems que puntúan en sentido contrario a la línea del test. A menudo, los ítems reversos contienen preguntas formuladas a través de negaciones. El empleo de este recurso (negaciones en las cuestiones que conforman los ítems) suele tener el sentido de evitar el efecto de atracción que ejercen determinados ítems, de modo que queden compensados respecto a estos. En el análisis del TEQ, se observa un descenso en la consistencia interna del test, existiendo en la escala cinco ítems formulados negativamente. Al eliminar los ítems negativos y efectuar la reducción de componentes sobre la nueva escala, compuesta por 11 ítems (tabla 4), se observa un coeficiente alfa de 0'79, menor que en la escala íntegra, compuesta por 16 ítems, aunque el análisis de componentes resulta mucho más satisfactorio al pasar de cuatro componentes poco diferenciados a tres componentes, que se constituyen en claros factores, cada uno de ellos diferenciado en relación a su cualidad o naturaleza. Los dos primeros obtienen una cierta suficiencia desde el punto de vista estadístico, quedando el tercero comprometido, al obtener un modesto coeficiente de Cronbach de 0'53.

Tal como se comentó, es requisito para considerar una prueba o test como unidimensional, que la medida de la fiabilidad sea alta. No obstante, una medida mediocre o baja en fiabilidad no indica necesariamente que no se pueda dar unidimensionalidad. Esto es lo que ocurre en nuestro tercer factor. A pesar del bajo índice obtenido en alfa de Cronbach, se observa una clara orientación en los ítems componentes hacia un tipo de conducta concreta y diferenciada respecto a los dos factores aislados.

A través de la división de la escala en dos subescalas (ítems directos / ítems inversos), se comprueba que los primeros disponen de mejores propiedades estadísticas. Esto quiere decir que posee una mejor consistencia interna ($\alpha = 0,82$ vs. $A = 0,72$), mejores cualidades para el análisis factorial, objetivadas a través de KMO y Chi-cuadrado en la prueba de esfericidad de Barlett.

Las diferencias cualitativas respecto a los factores que se representan a través de los componentes aislados (dos componentes en cada subescala) son patentes entre ítems directos e inversos, pero estas diferencias disminuyen en relación a los componentes que se dan dentro de cada subescala. De este modo, se obtiene que los ítems directos están vinculados al ámbito de la ayuda y conexión con los otros, mientras que los inversos están asociados a la falta de interés e indiferencia. Mientras que las directas no contienen ningún ítem formulado negativamente, las inversas contienen cinco de ocho.

La pérdida de calidad en las propiedades psicométricas en los ítems inversos habría que atribuirlos a la forma en que se plantean las preguntas o cuestiones. La consistencia interna de la prueba podría verse afectada al plantear preguntas que contienen negación al incrementarse la dificultad en la propia comprensión, por un lado, y por otro, aumentar la ambigüedad o el número de interpretaciones que el sujeto puede hacer.

En la práctica clínica, cuando se trabaja con cuestionarios, uno de los problemas o quejas más frecuentes por parte de los usuarios es, precisamente, la dificultad en la interpretación de los ítems formulados de esta manera.

Con la finalidad de explorar una escala que arroje un solo componente, se somete a análisis la escala compuesta por ítems directos, eliminando los elementos que son compartidos por las dos subescalas resultantes. Los ítems eliminados son los números 6 y 13, quedando una escala compuesta por seis elementos en lugar de ocho. Sometida a análisis factorial, se selecciona el ítem número 1, debido a, como ya se comentó en los resultados, que era el mejor candidato para ser eliminado, al obtenerse con ello una mejora en las propiedades psicométricas. De este modo, sí obtenemos una escala, ahora compuesta por cinco ítems, que ofrece un solo componente.

No obstante, el componente aislado es cualitativamente poco diferenciado respecto a la naturaleza del mismo, mezclándose áreas tales como aspectos emocionales asociados a vivencias ajenas (dos ítems), estar en sintonía emocional con otros (dos ítems) y conducta proactiva dirigida a la protección de los otros (un ítem).

Respecto a los datos observados por sexos, existen diferencias estadísticamente significativas entre hombres y mujeres. En la prueba original TEQ 16 ítems, se obtienen asociados a los cuartiles en puntuación directa, los siguientes resultados: Q1; 47; Q2: 52; Q3:55.

De acuerdo con el diagrama de porcentajes para los percentiles, las puntuaciones podrían considerarse en el siguiente sentido: nivel muy bajo, (hasta 39 puntos, es decir, percentil 5), nivel bajo (hasta 47 puntos, percentil 30), nivel medio (de 48 a 56 puntos, percentiles de 35 a 80) y nivel alto, (de 57 a 64 puntos, percentiles de 85 a 99). El percentil 50 cae en la puntuación con valor 52.

Tabla 4. Versión final TEQ 11. Ítems directos: 1, 2, 3, 4, 6, 7, 9 y 11; ítems inversos: 5, 8 y 10).

1. Cuando alguien se siente emocionado, yo también suelo emocionarme con él/ella, es como si sintiera lo mismo (por ejemplo, si alguien se pone triste, me pongo triste; si está contento/a, me pongo contento/a).
2. Me molesta ver que alguien está siendo tratado/a de manera irrespetuosa.
3. Disfruto haciendo que otras personas se sientan mejor.
4. Siento afecto y me preocupo por las personas menos afortunadas que yo.
5. Cuando un amigo o amiga empieza a hablar de sus problemas, trato de dirigir la conversación hacia otra cosa (cambiar de tema).
6. Puedo saber cuándo los demás están tristes incluso cuando no dicen nada.
7. Pienso que estoy “en sintonía” y con los estados de ánimo de otras personas, es decir, puedo sentirme como ellos/as se sienten.
8. Me enfado cuando alguien llora.
9. Siento un fuerte impulso de ayudar a una persona cuando veo que se encuentra mal o está molesta.
10. Me parece una tontería que la gente lllore por felicidad.
11. Me dan ganas de ayudar y proteger a una persona cuando veo que se aprovechan o quieren aprovecharse de él/ella.

Respecto a la muestra compuesta por hombres, los valores asociados a cuartiles son: Q1:44; Q2:49 Q3:53. En la muestra de mujeres, son: Q1: 49 Q2:53 Q3:56.

Conclusión

La reducción y explicación de los diferentes componentes que constituyen un constructo psicológico es uno de los objetivos de la psicometría. Por tanto, es en el marco del concepto teórico de unidimensionalidad donde se encontraría la mayor dificultad para desarrollar herramientas que midan de una forma concluyente características comunes respecto a lo que se quiere medir, diferentes de otras, bien sean similares o completamente diferentes. El problema de base es que tal unidimensionalidad podría no existir, al menos desde un punto de vista extremadamente reduccionista.

Precisamente, el TEQ se crea como un intento de unificar los diferentes puntos de vista teóricos que rodean al concepto de empatía. El creador del TEQ indica que si bien, el objetivo de la herramienta es crear una medida unitaria de empatía, no aspira a que ésta sea un recurso final, un medio definitivo, sino que se debe considerar como una herramienta más, complemento de las ya existentes.

El estudio realizado pretende aportar datos que ayuden en ese empeño. Después de que el TEQ fuese traducido y adaptado al idioma español, se procedió a su administración sobre una muestra lo más representativa posible de la población española. Tanto el estudio original de validación, como los realizados con posterioridad, adolecen de la limitación relativa a la cualidad de las características muestrales. Han sido realizados sobre colectivos concretos, por lo que su aplicabilidad quedaría restringida a efectos de inferencia y generalización.

Los primeros estudios de validación se realizaron en base a estudios correlacionales con otras pruebas, como el IRI, IPT-5, a través de análisis de validez convergente y discriminante, obteniendo como resultado una suficiente fiabilidad y validez, aislando un componente. Estudios posteriores realizados en diferentes lugares del mundo obtienen resultados similares en fiabilidad, incluyendo el presente estudio. En cambio, no existiría tanto acuerdo respecto a la medición y reducción de factores. En este sentido, los estudios de Corea encontraron tres factores. El estudio realizado en China sugiere una versión modificada y revisada para que fuera satisfactoria. Sin embargo, en otros estudios, se admite la suficiencia de la prueba para ser aplicada en las poblaciones equivalentes a aquellas sobre las que fue administrada.

En el presente estudio, se han identificado cuatro componentes que explican la conducta empática:

1. Sentir afecto, molestia por el malestar de otros y conducta de ayuda.
2. Indiferencia, falta de interés por el estado de los demás.

3. Asociado a sentir afecto y estar en sintonía con las emociones de los demás.
4. Indiferencia, falta de interés en los demás.

Cualitativamente, estos componentes parece que engloban tres dimensiones: los componentes 2 y 4 están claramente asociados a la falta de interés e indiferencia sobre los estados emocionales de los demás, mientras que el componente 1 parece vinculado a la conducta proactiva de ayuda, consecuencia del malestar debido a estados negativos de los otros. El componente 3 estaría más asociado a sentir el malestar ajeno.

Por otro lado, existirían diferencias entre hombres y mujeres en relación a la comprensión de las emociones del resto, obteniendo una mayor puntuación las mujeres. En principio, las mujeres tendrían una mayor aptitud para captar las emociones de los demás, implicarse en conductas de apoyo o asistencia. No obstante, se observa que también pueden mostrar indiferencia, falta de interés o molestia de signo emocional negativo (enfado, irritación) ante los estados emocionales negativos de los demás. En este componente la distancia entre medias, si bien es estadísticamente significativa, se reduce claramente respecto a los otros dos.

Dirigiendo el análisis hacia la reducción de los componentes, los resultados del EFA realizado bajo la condición de eliminar los ítems formulados de modo negativo, arrojan tres componentes diferenciados en el sentido comentado: el primero, asociado a la conducta de ayuda, el segundo, vinculado a compartir emociones y comprenderlas, y, finalmente, el tercero vinculado a falta de interés y comprensión de las conductas con carga emocional negativa en los otros.

Por último, la versión del cuestionario que ofrece una mayor fiabilidad junto con un número de componentes asociados a factores, es la compuesta por 11 ítems, versión seleccionada para ser validada en población española y empleada en próximas investigaciones.

Referencias

- Archer, D. y Costanzo, M. (1993). [The Interpersonal Perception Task-15 \(IPT-15\): A Guide for Researchers and Teachers \(Video-DVD\)](#). Berkeley: University of California.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. y Plumb, I. (2001). [The "Reading The Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-Functioning Autism](#). *Journal of Child Psychology and Psychiatry*, 42(2), 241-251 [DOI: 10.1111/1469-7610.00715].
- Baron-Cohen, S., y 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*

- tal Disorders*, 34(2), 163-175 [DOI: 10.1023/B:-JADD.0000022607.19833.00].
- Batson, C.D., Turk, C.L., Shaw, L.L. y Klein, T.R. (1995). Information function of empathic emotion: Learning that we value the other's welfare. *Journal of Personality and Social Psychology*, 68(2), 300-313 [DOI: 10.1037/0022-3514.68.2.300].
- Carranza Albán, B.L., y Huamani Urbina, N. (2020). [Adaptación del Cuestionario de Empatía de Toronto en jóvenes universitarios de Lima](#). Tesis de Licenciatura. Lima: Universidad César Vallejo.
- Davis, M.H. (1983). [Measuring individual differences in empathy: Evidence for a multidimensional approach](#). *Journal of Personality and Social Psychology*, 44(1), 113-126 [DOI: 10.1037/0022-3514.44.1.113].
- Fernández-Pinto, I., López-Pérez, B. y Márquez, M. (2008). [Empatía: Medidas, teorías y aplicaciones en revisión](#). *Anales de Psicología*, 24(2), 284-298.
- Hambleton, R.K. y Li, S. (2005). Translation and Adaptation Issues and Methods for Educational and Psychological Tests. En C.L. Frisby y C.R. Reynolds (Eds.), *Comprehensive Handbook of Multicultural School Psychology* (pp. 881-903). Nueva York: John Wiley & Sons.
- Hogan, R. (1969). Development of an Empathy Scale. *Journal of Consulting and Clinical Psychology*, 33(3), 307-316 [DOI: 10.1037/h0027580].
- Igartua Perosanz, J.J. y Paez Rovira, D. (1998). [Validez y fiabilidad de una escala de empatía e identificación con los personajes](#). *Psicothema*, 10(2), 423-436.
- Kourmoussi, N., Amanaki, E., Tzavara, C., Merakou, K., Barbouni, A. y Koutras, V. (2017). [The Toronto Empathy Questionnaire: Reliability and Validity in a Nationwide Sample of Greek Teachers](#). *Social Sciences*, 6(2), 62 [DOI: 10.3390/socsci6020062].
- Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S. y David, A.S. (2004). [Measuring empathy: reliability and validity of the Empathy Quotient](#). *Psychological Medicine*, 34(5), 911-924 [DOI: 10.1017/S0033291703001624].
- López-Pérez, B., Fernández-Pinto, I. y Abad García, F.J. (2008). *TECA: Test de empatía cognitiva y afectiva*. Madrid: TEA.
- Mehrabian, A. (1997). Relations among personality scales of aggression, violence, and empathy: Validation evidence bearing on the Risk of Eruptive Violence Scale. *Aggressive Behavior*, 23(6), 433-445 [DOI: 10.1002/(SICI)1098-2337(1997)23:6<433::AID-AB3>3.0.CO;2-H].
- Muñoz, J., Elosua, P., & Hambleton, R.K. (2013). [Directrices para la traducción y adaptación de los tests: segunda edición](#). *Psicothema*, 25(2), 151-157 [DOI: 10.7334/psicothema2013.24].
- Muñoz Leiva, P.S. y Valenzuela San Martín, K.P. (2016). [Validación preliminar Cuestionario de Empatía Toronto \(TEQ\) en una muestra de adolescentes chilenos de las regiones de La Araucanía y Los Ríos](#). Tesis de Maestría. Temuco (Chile): Universidad de la Frontera.
- Rankin, K.P., Kramer, J.H. y Miller, B.L. (2005). [Patterns of Cognitive and Emotional Empathy in Frontotemporal Lobar Degeneration](#). *Cognitive and Behavioral Neurology*, 18(1), 28-36 [DOI: 10.1097/01.wnn.0000152225.05377.ab].
- Seligman, M.E. (2004). *Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment*. Nueva York: Simon and Schuster.
- Spreng, R.N., McKinnon, M.C., Mar, R.A. y 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 [DOI: 10.1080/00223890802484381].
- Totan, T., Doğan, T. y Sapmaz, F. (2012). [The Toronto Empathy Questionnaire: Evaluation of Psychometric Properties Among Turkish University Students](#). *Eurasian Journal of Educational Research*, 46, 179-198.
- Xu, R.H., Wong, E.L., Lu, S.Y., Zhou, L., Chang, J. y Wang, D. (2020). [Validation of the Toronto Empathy Questionnaire \(TEQ\) Among Medical Students in China: Analyses Using Three Psychometric Methods](#). *Frontiers in Psychology*, 11, 810 [DOI: 10.3389/fpsyg.2020.00810].
- Yeo, S. y 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), 119 [DOI: 10.1186/s12909-021-02561-7].