

Evaluación de

Cuestionario de

Impacto del Trauma



RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	CIT. Cuestionario de Impacto del Trauma
Autor	María Crespo, Héctor González-Ordi, Mar Gómez-Gutiérrez y Pablo Santamaría
Autor de la adaptación española	--
Variable(s)	Psicopatología del estrés postraumático. Impacto y secuelas del padecimiento de acontecimientos traumáticos en adultos
Áreas de aplicación	Psicología clínica, Psicología Forense, Servicios sociales
Soporte	Lápiz y papel con corrección informatizada a través de la plataforma TEAcorrige

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Excelente	5
Fundamentación teórica	Excelente	5
Adaptación	N/A (No aplicable)	-
Análisis de ítems	Bueno	4
Validez: contenido	Excelente	5
Validez: relación con otras variables	Buena-Excelente	4.8
Validez: estructura interna	Buena-Excelente	4.5
Validez: análisis del DIF	Buena	4
Fiabilidad: equivalencia	N/A (No aplicable)	-
Fiabilidad: consistencia interna	Buena-Excelente	4.5
Fiabilidad: estabilidad	Buena	4
Fiabilidad: TRI	No se aporta	-
Fiabilidad inter-jueces	No se aporta	-
Baremos e interpretación de puntuaciones	Buena-Excelente	4.3

Comentarios generales

El Cuestionario de Impacto del Trauma (CIT) es un instrumento multidimensional, conformado por 180 ítems articulados en 30 subescalas que se dividen en cinco grandes bloques. El cuestionario está dirigido a evaluar el impacto vital del acontecimiento traumático sobre el estado psicológico y psicosocial de las personas afectadas. Su uso se ha concebido para contextos clínicos y médico/forenses.

Como puntos fuertes a destacar, cabría indicar que, frente a otros instrumentos disponibles que tienen como propósito también la evaluación del estrés postraumático, el CIT añade a la valoración de este trastorno la de otros aspectos psicológicos que son relevantes cuando se trata de analizar el impacto de un acontecimiento traumático. En concreto, la evaluación de la centralidad del acontecimiento, las escalas de síntomas y las escalas de alteraciones en las creencias, todos ellos elementos útiles para dicho propósito. En lo relativo a las escalas de síntomas, ha de reseñarse como aportación relevante la evaluación de indicadores de afectación psicológica de importancia en personas que han sufrido ciertos tipos de trauma (los provocados de modo intencional). En concreto, la dependencia emocional, el malestar sexual, las alteraciones en la conducta alimentaria, las conductas de riesgo y presencia de autolesiones e ideación/conducta suicida.

A destacar también las escalas de validez (inconsistencia y distorsión negativa) cuyo uso puede ser particularmente útil en el contexto médico/forense. La descripción de perfiles específicos, derivados de la descripción de las escalas y subescalas del cuestionario, es un elemento igualmente reseñable.

El proceso de construcción del cuestionario, amplia y detalladamente descrito en el manual, es un aspecto valorable. La revisión de las escalas e ítems del instrumento por parte de paneles de expertos, incluyendo a expertos de Latinoamérica para su posible uso futuro en otros contextos lingüísticos y culturales, es también un aspecto que destacar.

La fundamentación psicométrica es buena y bastante completa. Se ha llevado a cabo un gran esfuerzo por documentar el proceso de construcción del test. Presenta numerosas evidencias adecuadas, tanto de fiabilidad como de validez (contenido, validez convergente, referida a un criterio y estructura interna). Además, se trabaja con una muestra grande, que ha servido tanto para verificar las propiedades psicométricas como para construir los baremos, y muestras clínicas/forenses y de simuladores, que aportan evidencias de validez sobre las escalas de contenido y la de Distorsión negativa. Se ha procurado que el test sea comprensible y fácil de usar, con corrección automática a través de la plataforma online TEAcorrige.

En lo referente a posibles aspectos de mejora, cabe señalar que:

- La información básica como el número de ítems que forma cada escala es demasiado compleja de obtener (para algunas escalas no se ha logrado determinar el número de ítems). Es comprensible porque se trata de una batería formada por 30 escalas, pero no se reportan en muchos casos información a nivel de ítem (media y dispersión, discriminación - en algunos casos sí se indica el peso factorial obtenido en un análisis factorial confirmatorio).

- En cuanto al estudio de la estructura interna, sería necesario informar e interpretar las correlaciones entre los factores. También sería interesante llevar a cabo un estudio factorial conjunto de las escalas de los distintos bloques, lo que permitiría reforzar las conclusiones sobre la estructura interna.

- En cuanto al estudio de las propiedades psicométricas, sería ideal poder contar con varias muestras extraídas de la población general para estudiar la generalizabilidad de los resultados. En este sentido, parece que la versión obtenida de 180 ítems no se analizó con datos empíricos en un estudio piloto, sino que se consideró definitiva.

- Por último, el estudio del funcionamiento diferencial de los ítems (para las escalas de distorsión negativa) podría complementarse con un estudio de invarianza factorial del resto de las escalas, que permitiese verificar si la estructura puede asumirse a ser la misma para grupos relevantes según sexo y edad. Es esperable que este tipo de estudios se lleven a cabo en estudios de baremación posteriores.

- Aunque la interpretación se basa en las puntuaciones estandarizadas T, convendría revisar lo indicado sobre los percentiles, ya que se proporciona una única tabla, cuando en realidad cada escala debería tener su propia tabla.

En suma, el CIT es un instrumento útil para el profesional clínico y de especial valor para el ámbito de la evaluación clínica y forense.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

CIT. Cuestionario de Impacto del Trauma

1.2. Nombre del test en su versión original:

--

1.3. Autor/es del test original:

María Crespo, Héctor González-Ordi, Mar Gómez-Gutiérrez y Pablo Santamaría

1.4. Autor/es de la adaptación española:

--

1.5. Editor del test en su versión original:

Hogrefe TEA Ediciones

1.6. Editor de la adaptación española:

--

1.7. Fecha de publicación del test original:

2020

1.8. Fecha de publicación del test en su adaptación española:

--

1.9. Fecha de la última revisión del test:

2020

1.10. Área general de la/s variable/s que pretende medir el test:

Escalas clínicas

1.11. Breve descripción de la/s variable/s que pretende medir el test:

El propósito del Cuestionario de Impacto del Trauma (CIT) es realizar una evaluación de amplio espectro de los diversos síntomas asociados al padecimiento de un acontecimiento traumático en personas adultas (18-70 años de edad). Para ello, se efectúa una valoración global del impacto psicosocial de los eventos traumáticos, a fin de determinar cómo se configuran las secuelas psicológicas en la persona evaluada. La prueba puede ser empleada con cualquier individuo que haya sufrido un acontecimiento traumático de cualquier tipo para su uso en contextos aplicados de carácter clínico, forense y/o médico, así como del ámbito investigador. Puede ser respondida también en modalidad online y es posible obtener un informe interpretativo de sus puntuaciones

Se trata de una prueba de autoinforme, compuesta por un total de 180 ítems con un formato de cinco categorías de frecuencia (nunca o casi nunca, pocas veces, algunas veces, muchas veces, siempre o casi siempre). Para su uso autoaplicado, se requiere que la persona evaluada tenga un nivel de lectura similar a alumnado comprendido entre 4º y 6º de Educación Primaria. El tiempo de aplicación requerido es de entre 20 y 60 minutos, con un promedio de 45 minutos.

Los ítems se estructuran en 30 escalas dirigidas a la evaluación sintomatológica y que se agrupan en cuatro grandes bloques de contenido:

a) Bloque de estrés postraumático. Las escalas en este bloque permiten dilucidar, entre otras cosas, en qué grado están presentes los síntomas más relevantes y definitorios del estrés postraumático. Pertenecen a este bloque siete escalas:

- Intrusión.
- Evitación.
- Alteración cognitiva y del estado de ánimo.
- Alteración en la activación y reactividad.
- Disociación.
- Centralidad del acontecimiento.
- Puntuación total de estrés postraumático, que es un agregado de algunas de las previas (Intrusión, Evitación, Alteración cognitiva y del estado de ánimo y Alteración en la actividad y reactividad).

b) Bloque de sintomatología psicológica general. Mediante estas escalas se profundiza en qué áreas y dominios se manifiestan problemas y dificultades. En el área internalizante, pueden obtenerse puntuaciones en diez escalas:

- Problemas somáticos.
- Alteraciones del sueño.
- Ansiedad.
- Desregulación emocional.
- Dependencia emocional.
- Rumiación.
- Depresión.
- Desconexión social.
- Malestar sexual.
- Alteraciones de la conducta alimentaria.

En el área externalizante, pueden obtenerse puntuaciones en cinco escalas:

- Abuso de sustancias.
- Conductas de riesgo.
- Autolesiones.
- Ideación y conducta suicida.
- Ira.

c) Bloque de presencia de creencias negativas. Mediante estas escalas se valoran qué aspectos del sistema de creencias de la persona pueden estar afectados y deberían ser objeto de valoración o intervención. Se obtienen puntuaciones en cinco escalas:

- Creencias negativas sobre los demás.
- Creencias sobre el mundo y el futuro.
- Creencias sobre uno mismo.
- Creencias de culpa.
- Creencias de autosabotaje.

d) Bloque de deterioro en el funcionamiento, con una única escala. Mediante esta escala se valora en qué grado la persona evaluada percibe que la sintomatología está causando un deterioro significativo en lo social, laboral u otras áreas importantes de funcionamiento.

Además se incluye un bloque de escalas de validez. Mediante estas escalas se pretende identificar las limitaciones de las respuestas de la persona que completa el instrumento y valorar la adecuación o no de la información obtenida. Las escalas que forman parte de este bloque son:

- Omisiones
- Inconsistencia
- Distorsión negativa

El instrumento aporta una puntuación para cada escala que permite ser valorada de acuerdo a los baremos de referencia. Previa a la presentación de los ítems, se incluye un listado de 14 acontecimientos traumáticos (y la opción de describir algún otro no reflejado en el listado). La persona debe indicar el acontecimiento que le causa malestar en la actualidad, su gravedad y frecuencia, tiempo desde que sufrió el mismo por primera vez, la edad que tenía, las consecuencias del acontecimiento y si le ocurrió directamente o aconteció a otra persona..

1.12. Áreas de aplicación:

Psicología clínica

Psicología forense

Servicios sociales

1.13. Formato de los ítems:

Respuesta graduada / Tipo Likert (frecuencia: nunca o casi nunca, pocas veces, algunas veces, muchas veces, siempre o casi siempre).

1.14. Número de ítems:

180 en total. Atendiendo a las tablas proporcionadas, puede derivarse la siguiente información para cada escala (número de ítems entre paréntesis; ? = no se informa):

a) Bloque de escalas de estrés postraumático:

- Intrusión (5)
- Evitación (2)
- Alteración cognitiva y del estado de ánimo (7)
- Alteración en la actividad y reactividad (6)
- Disociación (?)
- Centralidad del acontecimiento (?)

La puntuación de Estrés postraumático total (20). Se obtiene a partir de las escalas anteriores (Intrusión, Evitación, Alteración cognitiva y del estado de ánimo y Alteración en la actividad y reactividad).

b) Bloque de escalas de síntomas:

- Problemas somáticos (10)
- Alteraciones del sueño (6)
- Ansiedad (8)
- Desregulación emocional (8)
- Dependencia emocional (7)
- Rumiación (4)
- Depresión (9)
- Desconexión social (6)
- Malestar sexual (6)
- Alteraciones de la conducta alimentaria (5)
- Abuso de sustancias (4)
- Conductas de riesgo (7)
- Autolesiones (5)
- Ideación y conducta suicida (6)
- Ira (7)

c) Bloque de escalas de alteraciones en las creencias:

- Creencias negativas sobre los demás (5)
- Creencias negativas sobre el mundo y el futuro (6)
- Creencias negativas sobre uno mismo (4)
- Creencias negativas de culpa (5)
- Autosabotaje (3)

d) Escala de deterioro en el funcionamiento(?)

Escalas de validez:

- Inconsistencia (30 pares de ítems ya incluidos en los 180)
- Distorsión negativa (?)

1.15. Soporte:

Papel y lápiz.

Informatizado.

1.16. Cualificación requerida para el uso del test de acuerdo con la documentación aportada:

Nivel C.

1.17. Descripción de las poblaciones a las que el test es aplicable:

Está pensado para la evaluación de adultos (18 años en adelante) de la población general o atendidos en contextos clínicos, forenses, sociales u otros, que hayan sufrido cualquier tipo de acontecimiento

traumático. Su funcionamiento ha sido evaluado en muestras de población general de adultos (18 años en adelante). Se ha ejecutado una revisión lingüística y cultural para que el contenido sea apropiado en muestra Latinoamericana. Se asume un nivel de lectura similar al alumnado de 4º-5º-6º de Educación Primaria. En el caso de no poseer este nivel (personas con analfabetismo funcional o con lengua materna no española), la prueba puede ser aplicada de modo extraordinario en formato oral por un/a evaluador/a con formación en Psicología o Psiquiatría.

1.18. Existencia o no de diferentes formas del test y sus características:

La batería presenta una única forma, aplicable en formato lápiz y papel o de manera online.

1.19. Procedimiento de corrección:

Automatizado por ordenador (plataforma de corrección online).

1.20. Puntuaciones:

Las puntuaciones se obtienen mediante la plataforma TEAcorrige empleando una clave de acceso (PIN).

a) Se procede a corregir la prueba si el número de respuestas en blanco o de doble respuesta está dentro del rango admitido (13 ítems o menos omitidos; 15% o menos de los ítems de la escala).

b) El sistema de corrección informatizada asigna automáticamente el valor promedio del informante al resto de ítems de esa misma escala cuando la respuesta está en blanco. No se informa si hay ítems inversos.

c) Para cada escala se ofrecen dos puntuaciones: una puntuación directa (promedio de respuestas a los ítems de una determinada escala) y su transformación en una puntuación en escala T (media = 50; desviación típica = 10), obtenida a partir de la comparación de la puntuación directa con el baremo de la población general (N = 1000). La puntuación de Estrés postraumático total se obtiene como promedio de las puntuaciones T en las escalas Intrusión, Evitación, Alteración cognitiva y del estado de ánimo y Alteración en la actividad y reactividad (p. 233), y se reescala para que tenga media = 50 y desviación típica = 10 (convendría explicitar esto último en futuras ediciones del manual).

d) Se ofrecen valores de puntuaciones T y su correspondencia con el percentil en la muestra de tipificación de población general del CIT, así como rangos de puntuaciones típicas de clasificación orientativa para interpretar las puntuaciones de las escalas de estrés postraumático, síntomas generales, creencias negativas y deterioro funcional. De acuerdo a ello, puntuaciones T con un valor comprendido en el rango 0-59 no tendrían significación aparente, puntuaciones con un valor de entre 60-69 indicarían posible significación clínica (presencia de sintomatología) y puntuaciones por encima de 70 indican elevación clínicamente significativa.

Para la puntuación de Estrés postraumático total se proporcionan otros puntos de corte, basados en las investigaciones realizadas: 62 o menos (no se presenta sintomatología traumática significativa), entre 63 y 73 (la persona presenta niveles significativos de sintomatología postraumática que pueden llegar a ser compatibles con TEPT), 74 o más (la persona presenta una sintomatología postraumática muy grave con una alta probabilidad de TEPT). Tal y como se indica en el manual, el primer punto de corte es más sensible pero menos específico y está asociado a una mayor tasa de falsos positivos (25%), mientras que el punto de corte de 74 implica una tasa de falsos positivos más reducida (8%).

e) La escala de inconsistencia para valorar si la persona ha respondido al azar o de modo inconsistente se puntúa en un rango que oscila entre 70 o menos (se interpreta que se ha respondido a la prueba de manera coherente y consistente) y 81 o más (indicativas de que los resultados no son válidos debido a un patrón de respuesta inconsistente que no permite interpretar el perfil de la prueba).

f) La escala de distorsión negativa permite valorar si la persona evaluada pretende proporcionar una distorsión particularmente negativa de sí misma, lo que afectaría a la interpretabilidad del resto de escalas. Puntuaciones iguales o por debajo de 82 indicarían que el protocolo es interpretable, entre 83 y 104 señalarían que es probable la distorsión negativa y sería necesario interpretar con cautela el resto de puntuaciones de la prueba, superiores a 105 indican alta probabilidad de distorsión negativa y cuestionarían la validez del protocolo.

1.21. Escalas utilizadas:

Puntuaciones directas

Puntuaciones estandarizadas T (media 50 y desviación típica 10).

Los percentiles no se utilizan en el informe de la prueba, si bien la tabla 3.1 recoge los correspondientes a las puntuaciones T 60, 65, 70, 75 y 80. No obstante, no está claro de dónde surgen estos valores orientativos, que, si se asocian a las distribuciones empíricas, serían distintos para cada escala.

1.22. Posibilidad de obtener informes automatizados:

Sí. En el manual aparecen ejemplos de informes y parecen completos y claros. Consta de varias páginas. El informe aporta, en la primera página, las características del acontecimiento traumático padecido (descripción de los eventos que la persona ha marcado como eventos padecidos), el acontecimiento que causa mayor nivel de malestar en la actualidad, la gravedad con que valora el mismo, si tuvo lugar de modo repetido, cuándo ocurrió por primera vez y qué edad tenía la persona, y si ha sido un acontecimiento que le ha ocurrido directamente o ha sido sufrido por otro individuo.

En la segunda página, se incluyen dos columnas con las puntuaciones en las 30 escalas (directas y escala T) acompañadas de un gráfico de líneas y puntos en el que representan las puntuaciones señalando horizontalmente T = 50 como línea de referencia y sombreando distintas regiones para facilitar la interpretación (p. ej., T en Inconsistencia menor que 70, entre 70 y 80, o superior a 80).

En la tercera página se incluyen comentarios cualitativos e información sobre los ítems críticos afirmados por la persona (ítems seleccionados para alertar rápidamente de posibles situaciones que requieren acción inmediata o la necesidad de solicitar clarificaciones).

Finalmente, en las últimas hojas se analiza el cumplimiento de los criterios diagnósticos DSM/CIE del trastorno de estrés posttraumático.

El informe, de carácter orientativo, se efectúa de modo automático en función de las respuestas más destacadas, por lo que su interpretación debe realizarse de forma conjunta con todos los datos adicionales provenientes de otras fuentes de información. En el mismo se detallan los aspectos fundamentales que caracterizan las respuestas de la persona a cada escala de acuerdo a su definición. Su redacción es clara y precisa, se centra en los aspectos fundamentales que definen cada subescala y aporta una noción bastante específica. Cabe destacar que efectúa recomendaciones de valoración adicional si la puntuación arrojada en algún apartado es particularmente elevada o en función de la respuesta a ciertas escalas. Finalmente, el informe aporta otras consideraciones orientativas derivadas del análisis conjunto de varias de las escalas.

1.23. Tiempo estimado para la aplicación del test:

En aplicación individual: 45 minutos aproximadamente (se indica un rango entre 20 y 60 minutos).

En aplicación colectiva: 45 minutos aproximadamente (se indica un rango entre 20 y 60 minutos).

1.24. Documentación aportada por el editor:

Manual. Cuadernillos, hojas de respuesta y PIN para 25 usos.

1.25. Precio de un juego completo de la prueba:

83,20 euros (IVA incluido) por un juego completo (Manual, 10 cuadernillos y Kit de corrección de 25 usos). Fecha de consulta: 25/11/21.

1.26. Precio y número de ejemplares del paquete de cuadernillos:

17,91 euros (IVA incluido) por un paquete de 10 cuadernillos.

Fecha de consulta: 25/11/21.

1.27. Precio y número de ejemplares del paquete de hojas de respuesta:

30,25 euros (IVA incluido) por Kit de corrección con 25 hojas de respuesta y PIN para 25 usos Fecha de consulta: 25/11/21.

1.28. Precio de la administración y/o corrección, y/o elaboración de informes por parte del editor:

6,05 euros (IVA incluido) por 1 uso de aplicación y corrección online y 12,10 euros (con IVA) por 1 uso de informe online. Fecha de consulta: 25/11/21.

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

2.1. Aspectos generales:

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	Excelente	5
2.2. Calidad de la documentación aportada	Excelente	5
2.3. Fundamentación teórica	Excelente	5
2.4. Adaptación del test	-	-
2.5. Desarrollo de los ítems del test	Excelente	5
2.6. Calidad de las instrucciones para el participante	Excelente	5
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Excelente	5
2.8. Facilidad para registrar las respuestas	Excelente	5
2.9. Bibliografía del manual	Excelente	5
2.10. Datos sobre el análisis de los ítems	Buenos	4

2.11. Validez:

2.11.1. Evidencias de validez de contenido:

Contenido	Valoración	Puntuación
2.11.1.1. Calidad de la representación del contenido o dominio	Excelente	5
2.11.1.2. Consultas a expertos	Se ha consultado a un amplio número de expertos mediante un procedimiento sistematizado (N > 30)	5

2.11.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y otras variables:

2.11.2.1. Evidencias de validez basadas en la relación entre las puntuaciones del test y otras variables:

Contenido	Valoración	Puntuación
2.11.2.1.1. Diseños empleados	Correlaciones con otros tests Diferencias entre grupos Diseños experimentales o cuasi experimentales	
2.11.2.1.2. Tamaño de las muestras	Varios estudios con muestras grandes (N > 500)	5
2.11.2.1.3. Procedimiento de selección de las muestras	<p>El muestreo fue incidental. La muestra de análisis es la misma muestra que se usa para construir los baremos. Con el fin de extraer una muestra representativa de la población general adulta, se selecciona a un total de 54 profesionales de varias regiones geográficas definidas (Norte, Sur, Este y Oeste de España). Se valoró la formación, titulación y experiencia de los profesionales. A cada profesional se le encargó evaluar a 12-43 casos según criterios establecidos de sexo y edad. Para fijar las variables de sexo, edad y región geográfica se atiende al Censo de Población y Viviendas del Instituto Nacional de Estadística. Se obtuvo una base de datos final de 1000 adultos. 247 de estos adultos fueron evaluados en una segunda ocasión para evaluar la estabilidad temporal (tras dejar pasar 28,3 días en promedio).</p> <p>Hay otra muestra clínico/forense empleada en la validación. Su selección se efectuó a través de un anuncio en la página web de la editorial donde la prueba ha sido publicada, en diversas redes profesionales y a través de la red de contacto de los autores. Los profesionales con interés completaron un formulacion online para conocer su formación, titulación, lugar de trabajo, experiencia profesional, contexto de evaluación y localización geográfica. De 110 solicitudes, se seleccionaron 65 solicitudes que cumplieron con los requisitos establecidos. Se llegó a una muestra final de 612 adultos (444 casos en contextos clínicos y 168 en contextos forenses).</p> <p>Finalmente, se emplean tres muestras de simuladores análogos donde siguiendo un procedimiento similar al de la muestra de la población general se llega a 471 adultos (muestra 1), 121 estudiantes de Psicología de la Universitat de les Illes Balears (muestra 2) y 61 asistentes a un</p>	

	<p>curso de verano sobre Psicología forense y detección de la simulación (muestra 3).</p> <p>De las muestras anteriores se derivan varias de las muestras para los estudios de validez convergente: una muestra de 612 adultos de la población general seleccionada al azar, de forma que estuviera emparejada en edad y sexo con la población clínica y forense, para las comparaciones con esta, así como distintas submuestras en las que se aplican pruebas específicas (p.ej., SIMS, EGEP-5, MMPI-2, etc.), con tamaño muestral variable.</p> <p>En términos generales se trabaja con muestras grandes, aunque algunos de los análisis se han efectuado con muestras de tamaño pequeño o moderado, debido a la segmentación por las características de las mismas o a que las escalas de validación han sido respondidas por subgrupos particulares. Tales casos pueden entenderse por el carácter clínico y forense de las muestras</p>	
2.11.2.1.4. Calidad de los tests empleados como criterio o marcador	Excelente	5
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares	Buena ($0.60 \leq r < 0.70$)	4
2.11.2.1.6. Promedio de las correlaciones con otros tests que miden constructos no relacionados	Buena	4
2.11.2.1.7. Resultados de la matriz multirasgo-multimétodo	No se aporta	-
2.11.2.1.8. Resultados de las diferencias intergrupo	Excelente	5

2.11.2.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y un criterio:

Contenido	Valoración	Puntuación
2.11.2.2.1. Criterios empleados	<p>Se compararon los diagnósticos de TEPT tomando como criterio la Escala Global de Estrés Postraumático 5; Crespo et al., 2017), aplicándose la misma a 460 personas de la muestra clínica y forense, de ambos sexos, diversas edades y procedencia geográfica.</p> <p>Para la interpretación de la escala de distorsión negativa se evaluó la capacidad de la misma para la detección de la simulación. Se comparan las puntuaciones de la muestra de población general, la muestra clínica y forense sin sospecha de simulación, la muestra clínica y forense con</p>	

	sospecha de simulación, y la muestra de simuladores análogos (instruidos para la simulación). Para establecer el subgrupo con sospecha de simulación se empleó el Inventario Estructurado de Simulación de Síntomas (SIMS; Widows y Smith, 2009).	
2.11.2.2.2. Calidad de los criterios empleados	Excelente	5
2.11.2.2.3. Relación temporal entre test y criterio	Concurrente	
2.11.2.2.4. Tamaño de las muestras	Varios estudios con muestras grandes o estudios meta-analíticos	5
2.11.2.2.5. Procedimiento de selección de las muestras	Incidental. Se efectuó mediante la colaboración de 65 profesionales clínicos y forenses que evaluaron a 444 casos en contextos clínicos y 168 en contextos forenses (véase respuesta abierta al epígrafe 2.11.2.1.3). Cabe destacar que se emplea un muestra clínico/forense grande.	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	Excelente ($r \geq 0.55$)	5

2.11.3. Evidencias de validez basadas en la estructura interna:

Contenido	Valoración	Puntuación
2.11.3.1. Resultados del análisis factorial	Buena-Excelente	4.5
2.11.3.2. Funcionamiento diferencial de los ítems	Buena	4

2.11.4. Acomodaciones en la administración del test:

Contenido	Valoración	Puntuación
2.11.4. El manual del test informa sobre las acomodaciones en la administración del test	<p>Sí. Se requiere que la persona sea capaz de leer y comprender las instrucciones de la prueba y los ítems que la componen. Si existen dudas sobre la capacidad lectora, se recomienda llevar a cabo una evaluación preliminar dirigida a determinar esto o al menos hacer algún tipo de cribaje. Se comenta que en circunstancias extraordinarias la prueba podría aplicarse en formato oral. Preferiblemente, debería grabarse un vídeo dictando los ítems de manera que la persona evaluada pueda responder con mayor autonomía e independencia. Finalmente, se comenta que en presencia de un problema de comprensión lectora es esperable su detección por la obtención de puntuaciones altas en la escala de validez de Inconsistencia.</p>	

2.11.5. Comentarios generales sobre evidencias de validez:

El manual del instrumento aporta amplia información relativa a la validez. Las evidencias se resumen en los siguientes puntos:

1. Evidencias de validez con relación al contenido. Las evidencias proporcionadas son muy satisfactorias, lo que garantiza una buena representación del dominio, de la adecuación de los ítems y de la calidad de las instrucciones. Se llegó a redactar un conjunto de 600 ítems, partiendo de definiciones detalladas de los constructos y sus facetas, basadas en una revisión de la literatura previa y de la sintomatología específicamente definida para el TEPT por los manuales DSM-5 y CIE-11. Destaca que se han empleado múltiples paneles de expertos para determinar la adecuación de los ítems, procedimientos sistematizados de valoración (p.ej., del grado de representatividad de las escalas en su conjunto, de la relevancia de cada una de ellas para la evaluación de personas que han sufrido un acontecimiento traumático, o de la claridad y relevancia de los ítems de las escalas para la evaluación del constructo evaluado). Además, se incluyó un panel de Latinoamérica para llevar a cabo una revisión lingüística y cultural y se hizo un análisis de la legibilidad de los ítems. También se utilizaron paneles de expertos para valorar los ítems de las escalas del bloque de validez (p. ej., valorando la frecuencia o prevalencia de los síntomas descritos).

2. Evidencias sobre la estructura interna de la prueba. Para el análisis de la estructura interna de la prueba, se ha llevado a cabo un análisis factorial confirmatorio de los ítems de las subescalas de estrés postraumático (intrusión, evitación, alteración cognitiva y del estado de ánimo, y alteración en la activación y reactividad) para establecer la pertinencia de cada ítem a cada una de ellas, efectuado tanto con los datos de la muestra de población general como con los de la muestra clínica, con resultados similares y satisfactorios en ambos casos. También se realizó un análisis factorial confirmatorio de las escalas de síntomas y de las escalas de alteraciones de las creencias, que mostraron un buen ajuste al modelo teórico en ambas muestras. Los procedimientos de estimación elegidos son adecuados a la naturaleza de los datos (p.ej. WLSMV para datos ordinales) y se proporciona información detallada de los resultados (i.e., saturaciones de los ítems en los factores e índices de ajuste).

3. Evidencias de validez de relaciones de las puntuaciones en el test y otras variables. Validez convergente con constructos similares. La información proporcionada es extensa, proporcionando correlaciones con las escalas del test Evaluación Global de Estrés Postraumático (EGEP-5 ; Crespo et al., 2017; N = 460), con las escalas del Inventario de Personalidad (PAI; Morey, 2011; N = 182), con la Escala de Experiencias Disociativas (DES-II; Carlson y Putnam, 1993; Icarán et al., 1996; N = 110), con la Escala de Centralidad del Evento (Berntsen y Rubin, 2006; N = 2204), con las escalas del MMPI-2 (Ben-Porath y Tellegen, 2009; N = 278), con el Test para la Evaluación de los Vínculos Interpersonales en Adultos (Vinculatest, Abuín, 2018; N = 110) y con las escalas del Listado de Síntomas Breve (LSB-50; Rivera y Abuín, 2012). Se emplea una muestra grande (N > 500) y otras de tamaño moderado o pequeño. Se constata una alta validez convergente (correlaciones en torno a 0,65). Por ejemplo, las correlaciones promedio de las escalas de Estrés postraumático con las del EGEP-5 es de 0,72; entre la puntuación de Estrés postraumático total con la subescala de estrés postraumático del PAI es de 0,75. En general la elección de las escalas es apropiada y se justifica adecuadamente, describiendo sus propiedades psicométricas.

Además, se incluyen comparaciones para las puntuaciones en escalas entre muestra de la población general (N = 612) y muestra clínica/forense, ya sea tomada globalmente (N = 612), o diferenciando por tipo de acontecimiento traumático sufrido (N entre 23 y 210), encontrándose tamaños del efecto generalmente medios o grandes, en correspondencia con las hipótesis formuladas.

4. Evidencias de validez de relaciones de las puntuaciones en el test con criterios externos. Se ha evaluado la capacidad diagnóstica de la prueba (sensibilidad y especificidad) mediante el uso de curvas ROC, tomando como criterio o "patrón oro" la clasificación proporcionada por la Escala Global del EGEP-5, en una muestra de tamaño moderado (N = 460). Los resultados arrojaron un valor satisfactorio del área bajo la curva (AUC = 0.86), indicativo de que la puntuación total de la escala de estrés postraumático del instrumento sometido a evaluación muestra adecuada capacidad para determinar el cumplimiento de los criterios diagnósticos del TEPT. La correlación con el criterio fue excelente (de 0,59). Los resultados acerca de la sensibilidad, especificidad y valor predictivo (positivo y negativo) de los puntos de corte propuestos son satisfactorios. Concretamente, un punto de corte de la prueba igual a 63 arroja una sensibilidad del 83%, y una tasa de falsos positivos del 25% (valor predictivo positivo: 86%; valor predictivo negativo: 71%), mientras que un punto de corte de 74 reduce la tasa de falsos positivos al 8%, manteniendo la sensibilidad en un 58% (valor predictivo positivo: 93%; valor predictivo negativo: 55%). Tomados en conjunto, estos datos arrojan evidencia acerca de validez de criterio adecuada en lo relativo al cumplimiento de los criterios diagnósticos. Los autores indican que los valores moderados de sensibilidad pueden deberse al diferente criterio para establecer la presencia de un síntoma (más estricto

en el caso del CIT).

Para las escalas de Inconsistencia y Distorsión negativa, también se recogieron evidencias de validez. Con respecto a la escala Inconsistencia, se incluye un estudio de simulación de respuestas aleatorias, a partir del cual se derivaron los puntos de corte (p.ej., con un 25% de respuestas aleatorias se detectaba un 70% de los casos con respuesta inconsistente). Con respecto a la escala de Distorsión negativa, se obtuvieron correlaciones con otras escalas sobre el mismo constructo (p.ej., el Inventario Estructurado de Simulación de Síntomas; SIMS, Widows y Smith, 2009; N = 717), encontrando resultados generalmente satisfactorios (p.ej., $r = 0,74$ para la muestra total). Asimismo, la escala de Distorsión negativa mostraba una elevada capacidad de discriminar las muestras de validación de las muestras de simuladores análogos. Por último, se consideró el funcionamiento diferencial de los ítems de la escala de Distorsión negativa con relación al sexo, y se descartaron ítems que mostrasen comportamiento diferencial.

Tan solo pueden valorarse algunos puntos de mejora:

- Aunque la información proporcionada es detallada, en análisis factorial es indispensable el reporte e interpretación de las correlaciones entre los factores. Igualmente, hubiera sido interesante un análisis factorial exploratorio conjunto de las escalas de distintos bloques, para poder valorar las relaciones entre las distintas dimensiones.

- Con relación al punto anterior, convendría concluir de manera más clara sobre la validez discriminante, resaltando qué correlaciones se espera que sean cero y constatando esto mismo empíricamente. En ocasiones se menciona, pero aparecen algunas correlaciones altas en casos donde los autores no indican que se trate de constructos relacionados.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	<p>Varios coeficientes de fiabilidad (para cada escala o subescala)</p> <p>Un único error típico de medida (para cada escala o subescala)</p> <p>Coeficientes de fiabilidad para diferentes grupos de personas</p> <p>Error típico de medida para diferentes grupos de personas</p> <p>Índice de cambio fiable</p>	

2.12.2. Equivalencia formas paralelas:

Contenido	Valoración	Puntuación
2.12.2.1. Tamaño de las muestras	-	-
2.12.2.2. Puesta a prueba de los supuestos de paralelismo	-	-
2.12.2.3. Promedio de coeficientes de equivalencia	-	-

2.12.3. Consistencia interna:

Contenido	Valoración	Puntuación
2.12.3.1. Tamaño de las muestras	Varios estudios con muestras grandes ($N \geq 500$)	5
2.12.3.2. Coeficientes de consistencia interna presentados	Coeficiente alfa o KR-20	
2.12.3.3. Promedio de coeficientes de consistencia	Buena ($0.80 \leq r < 0.85$)	4

2.12.4. Estabilidad (test-retest):

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	Un estudio con una muestra grande ($N \geq 200$)	3

2.12.4.2. Coeficientes de estabilidad	Excelente ($r \geq 0.80$)	5
--	-----------------------------	---

2.12.5. Cuantificación de la precisión mediante TRI:

Contenido	Valoración	Puntuación
2.12.5.1. Tamaño de las muestras	-	-
2.12.5.2. Coeficientes proporcionados	-	
2.12.5.3. Tamaño de los coeficientes	-	-

2.12.6. Fiabilidad inter-jueces:

Contenido	Valoración	Puntuación
2.12.6.1. Tipos de coeficientes presentados	-	
2.12.6.2. Promedio de los coeficientes	-	-

2.12.7. Comentarios generales sobre evidencias de fiabilidad:

Las evidencias de fiabilidad de las puntuaciones en la prueba sometida a valoración se han recogido en diversos estudios con muestras grandes, provenientes de la población general (N = 1000) y de una muestra clínica/forense (N = 612). A continuación, se resumen las evidencias obtenidas:

1. Consistencia interna. En lo que a la consistencia interna se refiere (alfa de Cronbach) su promedio es 0,81, habiéndose calculado en la muestra de la población general, en la muestra total clínico/forense, en la muestra clínica y en la muestra forense, por separado, y en función del sexo para todas estas muestras. El coeficiente promedio puede considerarse como bueno. El único valor bajo (0,66) se obtuvo para la escala de Autolesiones en la muestra de tipificación, lo que podría deberse a la menor variabilidad de la escala en esa muestra. Los coeficientes tienen valores ligeramente superiores para la muestra clínica/forense en todas las subescalas de la prueba, excepto en la subescala Rumiación. En algunos casos, se encuentran importantes diferencias de consistencia interna para las puntuaciones entre hombres y mujeres, aunque los autores indican que esas diferencias tienen que ver, generalmente, con la mayor o menor prevalencia del síntoma según el sexo, así como a la variabilidad asociada.

2. Coeficiente de fiabilidad test-retest. El valor promedio de los coeficientes de estabilidad temporal es de 0,86. Para su cálculo, se contó con una muestra de la población general de tamaño grande (N = 247), empleando un intervalo temporal de cuatro semanas. El valor más pequeño de los coeficientes de estabilidad temporal fue 0,75.

Además, se informa adecuadamente del error típico de medida y del Índice de cambio fiable. Esto último es particularmente útil para calcular el número de puntos de cambio necesarios para decidir con confianza qué sujetos han experimentado un cambio clínico relevante en la puntuación en la escala total. De acuerdo a dicho índice, una diferencia de 7 puntos sería significativa a un nivel de significación de 0,10, de 8 a 10 sería significativa a un nivel de 0,05, y de 11 o más puntos para un nivel de 0,01.

2.13. Baremación e interpretación de las puntuaciones:

2.13.1. Interpretación normativa de las puntuaciones:

Contenido	Valoración	Puntuación
2.13.1.1. Calidad de las normas	Un baremo adecuado para la población objetivo	3
2.13.1.2. Tamaño de las muestras	Muy grande ($N \geq 1000$)	5
2.13.1.3. Aplicación de estrategia de tipificación continua	No	
2.12.1.4. Procedimiento de selección de las muestras	<p>Incidental. Para la selección de la muestra de población general adulta (18-70 años) empleada para generar los baremos de población general, se selecciona a un total de 54 profesionales de varias regiones geográficas definidas (Norte, Sur, Este y Oeste de España). A cada profesional se le encargó evaluar a 12-43 casos según criterios establecidos de sexo y edad.</p> <p>Para conformar la muestra se siguieron las especificaciones y proporciones recogidas en el Censo de Población y Viviendas del Instituto Nacional de Estadística, controlando diversas variables para garantizar su representatividad, concretamente, edad, sexo y región geográfica. El número de casos para cada rango de dichas variables se determinó por los datos poblacionales establecidos en el mencionado censo. Además, se tuvo en cuenta la distribución porcentual del N total de la muestra de acuerdo con su estado civil, situación laboral y evento traumático padecido, a fin de valorar la variabilidad y representatividad de la muestra. No obstante, estos datos no fueron incluidos en la estratificación del muestreo.</p>	
2.12.1.5. Actualización de baremos	Excelente (menos de 10 años)	5

2.13.2. Interpretación referida a criterio:

Contenido	Valoración	Puntuación
2.13.2.1. Adecuación del establecimiento de los puntos de corte	Bueno	4
2.13.2.2. Procedimiento empleado para fijar	-	

los puntos de corte		
2.13.2.3. Procedimiento de obtención del acuerdo inter-jueces	-	
2.13.2.4. Valor del coeficiente de acuerdo inter-jueces	-	-

2.13.3. Comentarios generales sobre baremación e interpretación de las puntuaciones:

Se emplea una amplia muestra de baremación proveniente de la población general (N = 1000). Se cuidó en el diseño que la muestra fuese representativa de la población general adulta (entre 18 y 70 años) atendiendo al Censo de Población y Viviendas del Instituto Nacional de Estadística en cuanto a edad, sexo y región geográfica. Eso mismo se corroboró para otros datos como el estado civil y la situación laboral de las personas evaluadas. De cara a la interpretación, se reportan tanto puntuaciones directas como puntuaciones en escala T (media = 50; desviación típica = 10).

En el manual se ofrece una detallada información acerca de las normas de interpretación de la prueba. No obstante, también se advierte que ha de ser cada profesional quien adapte la interpretación del caso concreto que se halle evaluando. De hecho, se especifica que la interpretación de las puntuaciones y de los perfiles del instrumento tan solo debe ser realizada por profesionales cualificados y con formación específica en evaluación, psicopatología y estrés postraumático.

El perfil de resultados, que incluye la puntuación T para cada escala, se obtiene al comparar las puntuaciones directas de la persona evaluada con el baremo de la población general. Este método permite establecer una clasificación de las escalas de estrés postraumático, de síntomas, de alteraciones en las creencias y de deterioro, en función del rango de las puntuaciones T. Además, se pueden analizar los ítems críticos a partir del perfil facilitado, lo que es especialmente relevante para determinar el riesgo de suicidio, de autolesión o de conductas sexuales de riesgo o agresivas. Igualmente, dicho perfil facilita información acerca de los síntomas presentes para el diagnóstico del trastorno de estrés postraumático en cada una de las subescalas (intrusión, evitación, alteración cognitiva y del estado de ánimo, reactividad, disociación y centralidad del acontecimiento) y las pautas de interpretación de cada escala en función del rango de puntuaciones T.

Junto a la interpretación de la puntuación cuantitativa, se ofrece amplia información acerca de la interpretación de las puntuaciones elevadas en cada una de las subescalas clínicas de la prueba, aportando detalles tanto del significado de la escala como de la interpretación que se deriva de una elevada puntuación en la misma.

Con relación a los puntos de corte, para las puntuaciones en Estrés postraumático total (que se derivan de sumar las escalas de Intrusión, Evitación, Alteración cognitiva y del estado de ánimo y Alteración en la actividad y reactividad) se llevó a cabo una comparación con el diagnóstico realizado con el "patrón de oro" Escala Global de Estrés Postraumático 5. El análisis de curvas ROC indicó que área bajo la curva era 0,86. Analizando la curva ROC, se ponen los puntos de corte de $T > 62$ (sensibilidad 83% y especificidad 75%) y $T > 74$ (sensibilidad 58% y especificidad 92%). Para la escala de Distorsión negativa, siguiendo un procedimiento similar se consideraron las muestras clínicas y de simuladores para derivar los puntos de corte $T > 82$ y $T > 105$.

Los puntos de corte del resto de las escalas se basan en el criterio estadístico (i.e., una o dos desviaciones típicas por encima de la media) y están menos justificados. En concreto, se sugiere que puntuaciones T con un valor comprendido en el rango 0-59 no tendrían significación aparente, puntuaciones con un valor de entre 60-69 indicarían posible significación clínica (presencia de sintomatología) y puntuaciones por encima de 70 indican elevación clínicamente significativa. Sería conveniente reforzar la interpretación referida a criterio de esas puntuaciones.