

Evaluación de

Cuestionario TEA Clínico

- Revisado

∞CTC-R∞

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	CTC-R. Cuestionario TEA Clínico - Revisado
Autor	David Arribas, Sara Corral y Jaime Pereña
Autor de la adaptación española	-
Variable(s)	Sintomatología clínica y rasgos psicopatológicos
Áreas de aplicación	Psicología clínica Psicología forense Psicología del trabajo y las organizaciones Psicología del tráfico (EP) Clínicas de fertilidad
Soporte	Papel y lápiz Informatizado

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Excelente	5
Fundamentación teórica	Excelente	5
Adaptación	--	--
Análisis de ítems	Buena	4
Validez: contenido	Buena	4
Validez: relación con otras variables	Buena-Excelente	4.4
Validez: estructura interna	Buena-Excelente	4.5
Validez: análisis del DIF	Excelente	5
Fiabilidad: equivalencia	--	--
Fiabilidad: consistencia interna	Excelente	5
Fiabilidad: estabilidad	Adecuada-Buena	3.5
Fiabilidad: TRI	--	--
Fiabilidad inter-jueces	--	--
Baremos e interpretación de puntuaciones	Excelente	5

Comentarios generales

El CTC-R es un test de rendimiento típico de aplicación individual o colectiva destinado a evaluar sintomatología clínica y rasgos psicopatológicos en personas adultas. Consta de 110 ítems con un formato de respuesta Likert de 4 puntos. Incluye 10 escalas de distintos síntomas clínicos y características de personalidad que generan malestar, incapacidad o una respuesta desadaptativa y un índice de patología general (IPG), además de dos escalas de control de la validez (inconsistencia y deseabilidad social). Abarca una amplia variedad de escalas clínicas, sin que la aplicación sea excesivamente larga.

El CTC-R ofrece la opción de administrarse tanto mediante lápiz y papel como mediante internet, lo que se considera positivo desde el punto de vista de la accesibilidad. La información del manual sobre las instrucciones de aplicación de la prueba es muy detallada y clara. Los materiales resultan atractivos y bien estructurados para quien responde el test, aunque en soporte lápiz y papel el cuadernillo de preguntas y la hoja de respuestas se presentan de forma separada. El manual también aporta información muy detallada para ayudar a la interpretación de los resultados, incluyendo la posibilidad de disponer de un informe narrativo automatizado. La administración y corrección a través de la plataforma informática permite reducir los potenciales errores de corrección.

El manual aporta diferentes fuentes de evidencias de validez. En cuanto a las evidencias referidas al contenido, la experiencia de los autores, junto con el minucioso proceso de creación y depuración de los ítems, constituye una valiosa fuente de validez. Respecto a las evidencias basadas en la relación con variables externas, se muestran dos tipos de estudios. Primero, en cuanto a las evidencias convergente y discriminante, se valora positivamente que se aporten datos de siete estudios con diferentes instrumentos que evalúan personalidad, competencias o psicopatología, en muestras de diferente tamaño, aunque sería recomendable aportar algún dato sobre su calidad psicométrica en la muestra de cada estudio (p.ej., coeficiente de consistencia interna) e hipótesis claras sobre la dirección y magnitud de la correlación con las puntuaciones del CTC-R. Segundo, en los estudios de diferencias entre grupos, se utilizan muestras en las que "a priori" se ha asumido que el grado de deseabilidad social de los diferentes grupos en contextos competitivos o que pueden suponer un beneficio personal es superior al nivel estándar del grupo de población general. Convendría ampliar esta sección con la realización de estudios que relacionen las puntuaciones del CTC-R con criterios externos relevantes a nivel organizacional, clínico y/o forense (p.ej., rendimiento, diagnóstico, idoneidad, etc.). Las evidencias de validez basadas en la estructura interna del CTC-R dan soporte empírico al uso tanto de las puntuaciones de las 10 escalas clínicas como de la puntuación total (Índice de Patología General), y los resultados del análisis del funcionamiento diferencial de los ítems garantizan que la evaluación de hombres y mujeres está libre de sesgo, lo que es especialmente relevante en un proceso de selección de personal.

La fiabilidad se analiza desde la Teoría Clásica de Test y se proporcionan indicadores de la consistencia interna y la estabilidad temporal, con valores excelentes en ambos casos. Para la consistencia interna se utilizan muestras grandes segmentadas según el contexto de evaluación, mientras que para la estabilidad temporal se incluye una única muestra de población general de tamaño moderado. Se valora positivamente que se ofrezcan los valores del error típico de medida en escala T de la muestra de población general. Para ninguno de los indicadores de fiabilidad mencionados se incluyen los resultados de la puntuación total (IPG), aunque teniendo en cuenta los resultados del análisis factorial junto con los de consistencia interna y estabilidad de cada escala clínica, se prevén muy elevados.

Se valora favorablemente la existencia de 23 baremos en total, para 10 países diferentes, además de un baremo general para países de Latinoamérica. Para España, se dispone de 5 baremos en función del contexto de la evaluación (población general, selección de personal, cuerpos de seguridad, contexto forense, donación de gametos). Es destacable que las muestras utilizadas sean muy grandes, actuales, en su mayoría delimitadas con criterios de inclusión/exclusión, y en general representativas de la población objetivo, aun cuando se han obtenido de forma incidental. En base

a la premisa de que el diferente contexto de evaluación comporta diferentes grados de deseabilidad social en las respuestas, los 5 baremos se han establecido a partir de las diferencias observadas entre la población general y los cuatro restantes. Un punto fuerte es que las muestras de tipificación van actualizándose y, en consecuencia, se dispone de los datos vigentes correspondientes a cada momento de actualización (consultables a través del sistema informático TEACorrige).

En definitiva, las propiedades psicométricas del CTC-R son satisfactorias y avalan el uso previsto de sus puntuaciones.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

CTC-R. Cuestionario TEA Clínico - Revisado

1.2. Nombre del test en su versión original:

--

1.3. Autor/es del test original:

David Arribas, Sara Corral y Jaime Pereña

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:

2010 (CTC)

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

--

1.9. Fecha de la última revisión del test:

2022 (CTC-R)

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

Personalidad, Escalas clínicas

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

El CTC-R es un test que persigue evaluar la gravedad de distintos síntomas clínicos y características de personalidad que generan malestar, incapacidad o una respuesta desadaptativa en distintas esferas de la vida, como la personal o la laboral. Está destinado a población adulta. Puede utilizarse en diversos contextos, como son la práctica clínica y sanitaria, el peritaje forense y la selección de personal.

Consta de 10 escalas clínicas o de contenido, dos escalas para el control del sesgo de respuesta y un indicador general de patología.

Las escalas clínicas ofrecen un perfil detallado de la presencia e intensidad de la sintomatología clínica subyacente a determinadas afectaciones psicopatológicas. Estas escalas son: Preocupación por la salud (SAL), Ideación paranoide (PAR), Psicoticismo (PSI), Depresión (DEP), Límite (LIM), Hostilidad (HOS), Consumo de alcohol y sustancias (ALC), Impulsividad (IMP), Ansiedad (ANS) y Vulnerabilidad (VUL).

Por su lado, las escalas de control permiten valorar la validez del protocolo de evaluación, indicando, por un lado, si la persona ha prestado la suficiente atención a la hora de contestar o lo ha hecho de forma aleatoria (inconsistencia, INC) y, por otro lado, si ha tratado de ofrecer consciente o inconscientemente una imagen positiva o negativa sobre sí misma (deseabilidad social, DS).

Finalmente, el índice de patología general (IPG) ofrece un resumen de las escalas clínicas y proporciona una visión general sobre el grado de patología y afectación del individuo.

1.12. Áreas de aplicación:

Psicología clínica, Psicología forense, Psicología del trabajo y las organizaciones, Psicología del tráfico (EP), Clínicas de fertilidad.

1.13. Formato de los ítems:

Respuesta graduada / Tipo Likert

1.14. Número de ítems:

Total: 110 ítems (un mismo ítem puede contribuir a más de una escala)

Escalas clínicas: 96 ítems (distribuidos en las distintas escalas)

Escalas de control: 44 ítems

- Inconsistencia (INC): 30 ítems (15 pares de ítems entre los 96 anteriores)

- Deseabilidad social (DS): sin determinar.

Índice de patología general (IPG): 96 ítems de las 10 escalas clínicas, asumiendo que cada ítem contribuye a la puntuación de una única escala clínica.

1.15. Soporte:

Papel y lápiz, Informatizado.

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

Nivel B

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

El CTC-R es aplicable en población adulta. En el manual se indica que el uso de este instrumento resulta especialmente apropiado en el contexto clínico o terapéutico, en el ámbito de la psicología del trabajo y de las organizaciones, en el contexto forense, en las clínicas de fertilidad o en centros donde se valora la idoneidad de las personas para la conducción de vehículos, la renovación del permiso de conducir o la obtención del permiso de armas.

Aunque no existen requisitos específicos sobre el nivel educativo, en la descripción de la muestra de validación en población general española, se establece el siguiente criterio de inclusión: poseer el suficiente nivel de comprensión lectora y de competencia en español para comprender las instrucciones y el contenido de los ítems.

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

Si bien la prueba puede aplicarse mediante lápiz y papel y por internet, en ambos casos el cuestionario es idéntico, únicamente cambia la forma de administración. En el caso de la aplicación por internet, no existe ningún requisito inusual que se deba cumplir ni en lo referente al hardware ni al software; tan solo hay que entrar en la correspondiente plataforma de evaluación (TEAcorrige o e-teaediciones) y cumplimentar el cuestionario.

1.19. Procedimiento de corrección:

Automatizado por ordenador (plataforma de corrección on-line), y efectuada por la empresa suministradora (las hojas de respuesta se envían a la empresa para que esta se ocupe de la corrección).

1.20. Puntuaciones:

El CTC-R ofrece un total de 13 puntuaciones: una puntuación para cada una de las 10 escalas clínicas, una para el índice de patología general (IPG) y una puntuación para cada una de las 2 escalas de control. En cuanto a las escalas clínicas y el IPG, puntuaciones más elevadas son indicativas de mayor gravedad sintomatológica. En las escalas de control, puntuaciones más elevadas en la escala de inconsistencia (INC) indican mayor nivel de inconsistencia en las respuestas que ha dado la persona. Del mismo modo, en la escala de deseabilidad social (DS), puntuaciones más altas representan niveles más elevados de deseabilidad social.

La corrección de la prueba puede llevarse a cabo (a) mediante la plataforma de corrección de test por internet de Hogrefe TEA Ediciones (TEAcorrige) o (b) de forma mecanizada, en el que la persona a cargo de la evaluación envía las hojas de respuesta a Hogrefe TEA Ediciones. En ambos casos, es el propio sistema el que, según el baremo escogido, calcula automáticamente las puntuaciones (puntuaciones directas y puntuaciones T), ofrece el perfil de resultados y, opcionalmente, un informe narrativo. En este sentido, se desconoce la dirección de los ítems (directos o inversos), a qué escala pertenece cada ítem y cómo se obtiene la puntuación de cada escala (suma directa, suma ponderada, promedio, etc.).

Se ofrecen directrices precisas sobre el procedimiento a seguir cuando la persona ha anotado más de una respuesta a un mismo ítem (se anula el ítem) o ante la presencia de omisiones (para una omisión en una escala concreta, se reemplaza la omisión por el valor promedio de esa persona en la escala, y para más de una omisión en una escala concreta o más de 10 en el conjunto del test, se anula la escala o el test, respectivamente).

1.21. Escalas utilizadas:

T (Media 50 y desviación típica 10)

1.22. Posibilidad de obtener informes automatizados:

Sí. El sistema de corrección por internet (TEAcorrige) proporciona un perfil gráfico de resultados del CTC-R. En la cabecera del perfil se presentan los datos identificativos de la persona evaluada, información relativa a la sesión de evaluación y el baremo utilizado. En la parte central aparece la representación gráfica del perfil de puntuaciones, la cual permite resumir las puntuaciones directas y transformadas T de cada medida del CTC-R. El perfil utiliza un código de colores para representar los alejamientos relevantes respecto al punto central (T= 50). Finalmente, en la parte inferior se incluye un apartado de observaciones con las respuestas a los ítems críticos, o sea, aquellos que requieren de una atención inmediata por parte del profesional. Puede verse un ejemplo en: <https://web.teaediciones.com/Ejemplos/CTC-R-ejemplo-perfil.pdf>

También se puede disponer de un informe narrativo automatizado. Al inicio del informe se presentan

unas consideraciones previas que hay que tener en cuenta cuando se interpretan las puntuaciones del CTC-R de un individuo particular. Seguidamente, se presenta una interpretación cualitativa de los resultados de la persona evaluada divididos en tres secciones diferenciadas: (1) validez de la prueba, centrada en las escalas de control (INC y DS); (2) índice de patología general (IPG); y (3) perfil general en las escalas clínicas, donde la puntuación de cada escala se sitúa en una zona (estándar, alta o muy alta) según umbrales o criterios estadísticos preestablecidos.

1.23. Tiempo estimado para la aplicación del test:

En aplicación individual: 20-30 minutos.

En aplicación colectiva: 20-30 minutos.

1.24. Documentación aportada por el editor:

Manual. Cuadernillo, hoja de respuestas y clave de acceso para la aplicación y corrección por internet.

1.25. Precio de un juego completo de la prueba:

100,72€ (IVA incluido) por un juego completo (manual, 10 cuadernillos y kit de corrección de 25 usos). Fecha de consulta: 16/01/2023.

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

27,71€ (IVA incluido) por un paquete de 10 cuadernillos. Fecha de consulta: 16/01/2023.

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

43,55€ (IVA incluido) por el kit de corrección con 25 hojas de respuestas y 25 PINs de corrección. Fecha de consulta: 16/01/2023.

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

6,41€ (IVA incluido) por una aplicación y corrección online; 7,71€ IVA incluido) por el informe (que no incluye ni aplicación ni corrección del test); y 14,40€ IVA incluido) por una aplicación, corrección e informe online. Fecha de consulta: 16/01/2023.

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	Buena	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	Buena	4
2.11.1.2. Consultas a expertos	--	--

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 test Diferencias entre grupos	
2.11.2.1.2. Tamaño de las muestras	Varios estudios con muestras grandes	5
2.11.2.1.3. Procedimiento de selección de las muestras	<p>Incidental. En el estudio de las correlaciones con otros test de personalidad o psicopatología, se utilizan 7 muestras de conveniencia, mayormente provenientes de numerosos procesos de selección de personal, con un tamaño de 96 a 1468 participantes (promedio de 676).</p> <p>En el estudio de las diferencias entre grupos definidos por la finalidad de la evaluación, se emplean 5 muestras de 328 participantes cada una, seleccionadas aleatoriamente de las muestras de la tipificación española (descritas en el apartado 2.12.1.4), asegurando un equilibrio en el porcentaje de hombres y mujeres. También se incluye una muestra de conveniencia de 100 participantes provenientes de un centro de reconocimiento médico para la tramitación del permiso de conducción, con un 60% de hombres y un 40% de mujeres.</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	--	--

2.11.2.1.8. Resultados de las diferencias intergrupo	Buena	4
--	-------	---

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	--	--
2.11.2.2.2. Calidad de los criterios empleados	--	--
2.11.2.2.3. Relación temporal entre test y criterio	--	--
2.11.2.2.4. Tamaño de las muestras	--	--
2.11.2.2.5. Procedimiento de selección de las muestras	--	--
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	--	--

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	Excelente	5

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	No	

2.11.5. Comentarios generales sobre evidencias de validez:

A efectos de la validación del CTC-R, en el manual se describen detalladamente las características de las muestras provenientes de cada uno de los 5 contextos de evaluación: población general (n= 1008), selección de personal (n= 15214), cuerpos de seguridad (n = 11653), contexto forense (n= 454) y donantes de gametos (n= 1340).

El CTC-R es una versión revisada del CTC original. En el manual del CTC-R se presenta una descripción del proceso de construcción del CTC, que incluye (1) la delimitación de las facetas a incluir a partir de la revisión de diversos instrumentos ya existentes y de los manuales de trastornos mentales de referencia; (2) la redacción y depuración inicial de los ítems atendiendo a su relevancia, representatividad y claridad por parte de los autores del CTC; y (3) dos depuraciones sucesivas, ambas en base a criterios estadísticos (p.ej., índice de homogeneidad corregido), a partir de los datos recogidos en un primer estudio piloto con una muestra pequeña y otro estudio posterior con una muestra muy amplia de la población diana. La revisión del CTC que dio lugar al nuevo CTC-R se fundamentó en los siguientes aspectos: (1) se refinó el redactado de algunos ítems en base a la consulta a personas usuarias del CTC, para resolver problemas de vocabulario, comprensión, redacción, modismos, lenguaje ofensivo o elementos redundantes, llegando a eliminar algunos ítems problemáticos; (2) se eliminaron algunos ítems por presentar valores bajos de correlación ítem-total corregido; (3) se eliminó la escala de Neuroticismo por su solapamiento con las escalas de Depresión (DEP) y Ansiedad (ANS); (4) se renombraron algunas escalas para ajustarse mejor al contenido específico de sus ítems; (5) se modificaron las escalas de control, manteniendo una única escala de deseabilidad social (resultante de fusionar ítems de dos escalas anteriores) y creando una nueva escala de inconsistencia. Aunque no se ha consultado a especialistas externos acerca de la representación del contenido del CTC-R, la experiencia de los autores, junto con el minucioso proceso de creación y depuración de ítems, constituye una evidencia de validez referida al contenido.

Se aportan numerosas evidencias convergentes y discriminantes. Por lo que respecta al CTC-R, se proporcionan las correlaciones con 7 instrumentos que evalúan diferentes constructos: personalidad (16PF-5, NEO-PI-R, CPS, TOP, TPT), competencias (competeTEA) y psicopatología (PAI). Se emplean muestras de entre 96 y 1468 participantes (promedio 676) en su mayoría provenientes de diversos procesos de selección de personal.

Dadas las características de los instrumentos (todos multidimensionales) y la elevada cantidad de instrumentos utilizados en el proceso de validación, es complicado calcular el promedio de las correlaciones del CTC-R con otros test, tanto con aquellos que miden el mismo constructo o constructos similares como con otros test que miden constructos con los que el CTC-R no debería estar relacionado, y valorar la adecuación de estas correlaciones en el contexto de un análisis de evidencias convergentes y discriminantes. Sin embargo, salvo algunas pocas excepciones, se muestran correlaciones moderadas y altas con los constructos con los que se espera que haya correlación y correlaciones bajas con los constructos con los que no se espera que haya relación. Por lo que se puede concluir que existen evidencias convergentes y discriminantes que apoyarían el uso de las puntuaciones del CTC-R.

Además de las correlaciones del CTC-R con otros instrumentos, en el manual se aportan los resultados del análisis de las diferencias en el CTC-R en función del contexto de evaluación. Concretamente, se compara las puntuaciones de la población general con cuatro grupos provenientes de contextos competitivos y/o que implican un beneficio personal (selección de personal, cuerpos de seguridad, contexto forense y donación de gametos). Cada una de las submuestras utilizadas tiene un tamaño moderado (328 participantes por grupo), cuyos casos se seleccionaron aleatoriamente de las 5 muestras de validación equilibrando el porcentaje de casos de cada sexo dentro de cada muestra. Con tamaños del efecto entre medianos y grandes (g de Hedges entre 0.65 y 1.16), los resultados muestran que existen niveles más elevados de sintomatología en la población general en comparación con los demás contextos. Los autores del CTC-R atribuyen estas diferencias al hecho de que, en contextos competitivos y/o que suponen un beneficio personal, las personas tienden a infraestimar su

sintomatología clínica al pensar en las posibles consecuencias que puedan tener (p.ej., conseguir o no un puesto de trabajo). Esta explicación concuerda con las puntuaciones obtenidas en la escala de deseabilidad social (DS), donde la población general es la muestra con puntuaciones más bajas. Estos resultados, en conjunto, apoyan la utilización de 5 baremos diferenciados en función del contexto de evaluación y aportan evidencias de validez a favor del uso del CTC-R. Finalmente, también se muestra la comparación entre las puntuaciones T que resultan de aplicar cada uno de los 5 baremos en una muestra de conveniencia de 100 personas en un contexto de tramitación del permiso de conducir, que lleva a plantear que el baremo más conveniente para este grupo es el del contexto forense.

Por último, en el caso de las dos escalas de control, se aportan algunas evidencias de validez. Para la escala de inconsistencia (INC), se establece un punto de corte a partir de la sensibilidad y la especificidad obtenidas al comparar las puntuaciones T observadas en los grupos de población general y de selección de personal con puntuaciones generadas aleatoriamente. Y para la escala de deseabilidad social (DS), se evalúa si la puntuación discrimina entre los diferentes grupos según el contexto de evaluación y, aunque no son variables externas al test, se obtiene la correlación con cada escala clínica y con el índice de psicopatología general (IPG).

Sobre las evidencias basadas en las relaciones entre las puntuaciones del CTC-R y una variable criterio, sería interesante diseñar estudios con criterios externos relevantes a nivel organizacional, clínico y/o forense (p.ej., rendimiento, diagnóstico, idoneidad, etc.).

Se aportan evidencias de la estructura interna del test a través de la aplicación del análisis factorial exploratorio (AFE) con dos aproximaciones diferentes. Por un lado, además de la obtención del índice de discriminación mediante la correlación ítem-total corregido, el análisis de ítems plantea un AFE unidimensional de cada una de las 10 escalas clínicas en las 5 muestras según el contexto de aplicación. Se utiliza el estimador WLSMV, acorde con el formato de respuesta de los ítems. Los resultados a nivel de magnitud de las saturaciones factoriales e índices de ajuste son satisfactorios prácticamente en la totalidad de los casos. Por otro lado, se presentan los resultados de varios AFEs, uno por cada muestra de validación, con el método de estimación ULS acorde con la falta de normalidad de la distribución de las medidas, usando las puntuaciones de las 10 escalas clínicas como variables observadas. Aunque no se menciona que se hayan verificado las condiciones de aplicación (medida KMO y/o prueba de esfericidad de Bartlett), la matriz de intercorrelaciones entre escalas que se proporciona para dos de las muestras, con valores mayormente altos, parece indicar que es pertinente realizar el análisis factorial. Como criterio para determinar el número de dimensiones se ha usado una variante del análisis paralelo. Todos los análisis aconsejaron una dimensión general a extraer. El porcentaje de varianza explicada por la solución unidimensional (53-73%), junto con las elevadas cargas factoriales, apoyaría la existencia de un único factor de segundo orden: el índice de patología general (IPG). Adicionalmente, se estudia el funcionamiento diferencial de los ítems (DIF, de sus siglas en inglés) (comunicación personal, 2023) tomando el sexo como variable de agrupación (varones vs. mujeres). El DIF se estudia separadamente en la muestra de población general y la muestra de selección de personal, ambas de tamaño muestral elevado. Se aplican los procedimientos POLYSIBTEST, regresión logística ordinal y Mantel, y se obtienen las correspondientes medidas del tamaño del efecto. Los resultados son muy satisfactorios, encontrándose un porcentaje muy bajo de ítems con DIF en ambas muestras (4.5% y 0.9%, respectivamente, para la muestra de población general y de selección de personal).

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Varios coeficientes de fiabilidad (para cada escala o subescala) Un único error típico de medida (para cada escala o subescala) Coeficientes de fiabilidad para diferentes grupos de personas	

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	5
2.12.3.2. Coeficientes de consistencia interna presentados	Alfa ordinal	
2.12.3.3. Promedio de coeficientes de consistencia	Excelente ($r \geq 0.85$)	5

2.12.4. Estabilidad (test-retest):

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	Un estudio con una muestra moderada ($100 \leq N < 200$) o varios estudios con muestras pequeñas ($N < 100$)	2
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:

El estudio de la fiabilidad de las puntuaciones del CTC-R se centra en la consistencia interna y la estabilidad temporal de las 10 escalas clínicas y de la escala de discapacidad social (DS).

La consistencia interna se evalúa en las diferentes muestras (población general, selección de personal, cuerpos de seguridad, contexto forense y donantes de gametos) con un tamaño muestral grande (entre 454 y 11653 participantes). Se emplea el coeficiente alfa ordinal, dado el formato de respuesta de los ítems que consiste en una escala Likert de 4 puntos. Los valores obtenidos muestran una consistencia interna excelente, ya que el valor promedio de todos los coeficientes es de 0.86. En cada muestra, los valores obtenidos pueden considerarse entre buenos y excelentes, dado que los valores promedio oscilan entre 0.81 y 0.91. Sería deseable conocer el valor del coeficiente alfa ordinal del Índice de psicopatología general (IPG), aunque teniendo en cuenta los resultados del análisis factorial junto con los de consistencia interna, se augura que será muy elevado. Se concluye que la consistencia interna de las escalas del CTC-R es muy satisfactoria.

Se valora positivamente que, utilizando los coeficientes de consistencia interna de la muestra de la población general, se ha calculado para cada escala el error típico de medida (ETM) en puntuaciones T, así como el error máximo (EM) a los niveles de confianza del 90%, 95% y 99%.

Los indicadores de estabilidad temporal de las puntuaciones T, con un intervalo de 2 a 4 semanas, son muy satisfactorios. Para el análisis, se emplea una única muestra de población general, de tamaño moderado (130 participantes), pero que incluye un amplio rango de edad y está equilibrada por sexo. Se obtienen dos indicadores. Por un lado, el valor promedio del coeficiente de correlación de Pearson es de 0.84 (valores por escalas entre 0.75 y 0.89). Y, por otro lado, la comparación entre las dos puntuaciones T en el test y el retest indican que no existen diferencias estadísticamente significativas ni en las 10 escalas clínicas ($p \geq 0.131$) ni en la escala discapacidad social ($p = 0.011$, nivel crítico de alfa corregido = 0.004), con un tamaño del efecto nulo (g de Hedges ≤ 0.06 y $g = 0.12$, respectivamente). De todo esto se concluye que la estabilidad temporal de las puntuaciones obtenidas mediante el CTC-R es excelente.

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	Amplio rango de baremos en función de la edad, el sexo, el nivel cultural y otras características relevantes	5
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 elaboración de los baremos españoles, se ha utilizado un amplio número de muestras incidentales, las cuales se diferencian por el contexto de aplicación. En todas ellas se ha procurado una amplia representación de distintos aspectos sociodemográficos, en algunos casos a partir de una selección aleatoria respecto a muestras originales más grandes. En general, se observa un amplio rango de edad y nivel de estudios, y, en algunos casos, incluso equilibrio por género y por zona geográfica atendiendo a las diferentes comunidades autónomas.</p> <p>Muestra de población general (n= 1008): La muestra de población general se ha recabado mediante un procedimiento no aleatorio, pero muy sistematizado. Las cuotas se han establecido teniendo en cuenta el censo de población española con relación a tres variables: zona geográfica (Norte, Sur, Este y Centro), edad (6 grupos de edad) y sexo (50% mujeres y 50% varones). Se definieron diversos criterios de exclusión para garantizar que se tratara de un contexto no competitivo, con participantes sin afectación psicológica diagnosticada y libres de tratamiento por un trastorno mental. Para el reclutamiento se contó con la colaboración de 29 profesionales.</p> <p>Muestra de selección de personal (n= 3539): Se partió de una muestra de 11653 casos (88% hombres) evaluados con fines de selección de personal. Con la finalidad de equilibrar la muestra por sexo, se eliminaron al azar numerosos participantes, lo que supuso una muestra final de 3539 (50% hombres) procedentes de 173 procesos de selección diferentes.</p> <p>Muestra de cuerpos de seguridad (n= 11653): No se especifica el procedimiento de selección de la muestra de cuerpos de seguridad. Se indica que las evaluaciones se llevaron a cabo en un contexto competitivo de selección de personal u oposición pública y se describen las principales</p>	

	<p>características sociodemográficas de la muestra (Media edad= 37.2, DT = 7.9; 91% hombres).</p> <p>Muestra del contexto forense (n= 454): Para recopilar la muestra evaluada en un contexto forense, se contó con la colaboración de 61 profesionales que incluyeron durante unos meses el CTC entre sus herramientas de peritaje. A pesar de que no se especifica el procedimiento de selección de la muestra, se indica que se descartaron los casos en los que la evaluación tuviera la finalidad de determinar una lesión o secuela psíquica. Se describen las principales características sociodemográficas de la muestra (Media edad= 40.0, DT= 7.9; 64% hombres; 49% sin estudios).</p> <p>Muestra de donantes de gametos (n= 1340): No se especifica el procedimiento de selección de la muestra. Se describen las principales características demográficas (Media edad= 24.6, DT= 4.5; 86% mujeres).</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	--	--
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:

Los baremos del CTC-R se van actualizando (última actualización en el momento de elaborar este informe: 2022), con lo que se dispone de información de las muestras de tipificación vigentes en la plataforma de corrección informática (www.teacorrige.com). A continuación, se presenta información relativa a los baremos descritos en el manual.

La interpretación de las puntuaciones es normativa. A partir de 55016 evaluaciones, se proporcionan 23 baremos diferentes, en función del país (10 en total, en todos los casos de habla hispana) y de la finalidad de la evaluación (variable según el país, de 1 hasta 5 diferentes por país). Las muestras utilizadas para la baremación incluyen entre 202 y 11653 participantes y, en promedio, son muy grandes (promedio por baremo de más de 2000 participantes).

En el caso de España, se dispone de 5 baremos diferenciados según la finalidad de la evaluación, tanto en contextos competitivos (selección de personal, cuerpos de seguridad) o que pueden suponer la obtención de cierta recompensa o beneficio personal (forense,

donación de gametos), como no competitivos (población general). Las muestras utilizadas oscilan entre 454 y 11653 participantes y, en promedio, son muy grandes (promedio por baremo de casi 3600 participantes). Si bien el muestreo no ha sido aleatorio, en la composición de la mayoría de las muestras se ha procurado que queden representadas diferentes características sociodemográficas relevantes (edad, nivel educativo, sexo y, para la muestra de población general, también la zona geográfica a partir de datos censales). Los autores justifican los diferentes baremos en base a las diferencias de las puntuaciones entre los grupos considerados según la finalidad de la evaluación.

Los baremos ofrecidos a través de la plataforma de corrección online transforman las puntuaciones directas en puntuaciones T (media 50 y desviación estándar 10). Además, en el manual se detalla la correspondencia entre puntuaciones T y puntuaciones z y otras puntuaciones estandarizadas derivadas (S, decatipo, eneatipo) y se advierte que no se proporciona la correspondencia a percentiles dada la distribución de las medidas. A nivel de descriptivos, en el manual se proporcionan los indicadores de asimetría, curtosis y la significación de la prueba de normalidad de las 10 escalas clínicas y las 2 escalas de control para cada baremo/grupo, pero no son accesibles los valores de la media y la desviación estándar, ni en el propio manual ni a través de la plataforma digital.

En el manual se proporcionan pautas de interpretación de las puntuaciones T, estableciéndose niveles de rasgo en base a un criterio estadístico (alejamiento de la media): 20-59 "estándar", 60-69 "en riesgo" y ≥ 70 "muy alto" para las 10 escalas clínicas y el Índice de Psicopatología General (IPG). En el caso de la escala de control inconsistencia (INC) se ofrece una pauta de interpretación utilizando los mismos 3 niveles que en las escalas clínicas, mientras que para la escala de control deseabilidad social (DS) se presentan 5 niveles junto con su pauta de interpretación. En el informe narrativo automatizado, estos umbrales de nivel de rasgo son utilizados como zonas para describir el perfil de las diferentes escalas y se corresponden con los diferentes colores del perfil gráfico que proporciona la corrección automática.