

Guillem Feixas<sup>1,2</sup>  
Eric Badia<sup>1</sup>  
Arturo Bados<sup>1</sup>  
Joan C. Medina<sup>1,2</sup>  
Antoni Grau<sup>3</sup>  
Ernesto Magallón<sup>1</sup>  
Luis Botella<sup>4</sup>  
Chris Evans<sup>5</sup>

# Adaptación y propiedades psicométricas de la versión española del YP-CORE (*Young Person's Clinical Outcomes in Routine Evaluation*)

<sup>1</sup>Departamento de Psicología Clínica y Psicobiología, Universitat de Barcelona

<sup>2</sup>Institut de Neurociències, Universitat de Barcelona

<sup>3</sup>Institut de Trastorns Alimentaris, Barcelona

<sup>4</sup>Facultat de Psicologia, Ciències de l'Educació i l'Esport Blanquerna, Universitat Ramon Llull, Barcelona

<sup>5</sup>Psychology Department, University of Roehampton, Londres

**Introducción.** Ante la creciente prevalencia de problemas de salud mental en la población general, resulta imprescindible el uso de instrumentos de evaluación destinados a evaluar el resultado de las intervenciones terapéuticas, con el fin de facilitar el proceso de rehabilitación psicológica.

**Metodología.** Se expone el proceso de adaptación al español y un primer estudio psicométrico del *Young Person's Clinical Outcomes in Routine Evaluation* (YP-CORE), un instrumento diseñado para medir el resultado de las intervenciones terapéuticas a través de la evaluación del malestar general en población infanto-juvenil (11-16 años). Para ello, se ha tenido en cuenta una muestra clínica de 104 participantes y otra no clínica de 131.

**Resultados.** Los análisis han demostrado una buena aceptabilidad, una consistencia interna adecuada, así como una estabilidad test-retest aceptable, con correlaciones moderadamente elevadas. Además, el instrumento ha reflejado correlaciones significativas con todas las dimensiones del *Youth Self Report*, siendo las más elevadas entre ambas puntuaciones totales. También resulta útil a la hora de discriminar entre la población clínica y no clínica, mostrando poca influencia de la edad, pero sí del sexo, con puntuaciones más elevadas en el género femenino. El Análisis de Componentes Principales replica la estructura original. También se proporcionan los puntos de corte que servirán para calcular el cambio fiable y clínicamente significativo.

**Conclusiones.** Los resultados expuestos permiten hacer un uso fiable del instrumento, aunque se mencionan ciertas limitaciones que sugieren la necesidad de realizar investigaciones con muestras mayores y más representativas, en las cuales se verifiquen las propiedades psicométricas del instrumento.

**Palabras clave:** YP-CORE, Adaptación de test, Español, Adolescentes, Propiedades psicométricas

*Actas Esp Psiquiatr* 2018;46(3):75-82

## Adaptation and psychometric properties of the spanish version of the YP-CORE (Young Person's Clinical Outcomes in Routine Evaluation)

**Introduction.** Given the increasing prevalence of mental health problems in the general population, it is indispensable to use assessment tools aimed to assess the outcome of therapeutic interventions in order to refine the process of psychological rehabilitation.

**Method.** We describe the process of adaptation into Spanish and a first psychometric study of the *Young Person's Clinical Outcomes in Routine Evaluation* (YP-CORE), an instrument designed to measure the outcome in terms of general distress of therapeutic interventions in young people (11-16 years). 104 adolescents participated in the clinical and 131 in the non-clinical samples.

**Results.** Analyses showed good levels of acceptability, adequate internal consistency and acceptable test-retest stability, with moderately high correlations between administrations. In addition, the instrument yielded significant correlations with all dimensions of the *Youth Self Report*, the highest being between both total scores. Crucially, discriminated between clinical and non-clinical samples and showed a small effect of age but a larger effect of gender, with higher scores for females. The Principal Component Analysis replicates the original structure. Cut-off scores to calculate the reliable and clinically significant change are provided.

**Conclusions.** These results support initial use of the instrument though there are certain limitations that indicate the need for more research with larger and more represen-

Correspondencia:

Guillem Feixas

Departamento de Psicología Clínica y Psicobiología, Facultad de Psicología, Universitat de Barcelona

Passeig Vall d'Hebron, 171, 08035 Barcelona, España

Tel.: (+34) 933125100

Correo electrónico: gfeixas@ub.edu

tative samples, in which the psychometric properties of the instrument should be verified.

Keywords: YP-CORE, Test adaptation, Young people, Spanish, Psychometric properties

## INTRODUCCIÓN

Existe evidencia empírica acerca de la creciente prevalencia de los problemas en salud mental entre la población infanto-juvenil<sup>1</sup>, con tasas que se sitúan entre el 17 y el 20%<sup>2</sup>. Para abordar esta problemática resultan particularmente relevantes aquellas estrategias que permitan mejorar la atención en salud mental de esta población, especialmente los programas de tratamiento para hacer frente al amplio abanico de dificultades que puede presentar<sup>3</sup>. Concretamente, es esencial aplicar herramientas y tratamientos eficaces y útiles clínicamente<sup>4</sup>, que incidan sobre los aspectos que provocan un mayor deterioro bio-psico-social, procurando fomentar una mejora sustancial en el estado mental y la calidad de vida de los adolescentes.

En este sentido también es importante monitorizar los cambios producidos a través del proceso terapéutico, lo cual puede ayudar a obtener una medida más exhaustiva del progreso y su contraste entre enfoques, centros y profesionales<sup>5</sup>. Para ello, comúnmente se han utilizado distintas herramientas, de entre las que destacan la observación directa, la entrevista, los autorregistros y, fundamentalmente en población infanto-juvenil, las técnicas proyectivas<sup>6</sup>.

A pesar de ello, las técnicas citadas pueden resultar insuficientes, por lo que es crucial el uso de instrumentos de evaluación que faciliten el registro de los cambios producidos de forma cuantitativa y permitan detectar la necesidad de modificar el curso de la intervención terapéutica. Muchos de los desarrollados hasta el momento se centran en medir la sintomatología de cada patología en particular, siendo escasamente aplicables a pacientes con otras problemáticas o con diagnósticos inciertos o poco definidos<sup>7</sup>, algo relativamente frecuente en la adolescencia. Así, existen pocas medidas validadas para la franja de edad infanto-juvenil, la mayoría centrándose en casos cuya severidad es muy elevada<sup>8</sup>. De entre las más significativas y utilizadas actualmente pueden citarse, entre otros, el *Children Depression Inventory*/Inventario de Depresión para niños<sup>9</sup> (CDI), el Cuestionario de Ansiedad Estado/Rasgo para niños<sup>10</sup> (STAIC), el *Strengths and Difficulties Questionnaire*/Cuestionario de Fortalezas y Debilidades<sup>11</sup> (SDQ), o el *Youth Self Report*<sup>12</sup> (YSR) derivado del *Child Behaviour Checklist*/Inventario de comportamiento en niños<sup>13</sup> (CBCL), y del cual se utilizará la versión en español de Achenbach y Rescorla<sup>14</sup> en el presente estudio.

Todo ello apoya la necesidad de elaborar instrumentos breves que eviten el uso de múltiples instrumentos y recursos<sup>15</sup>, fomenten la evaluación cuantitativa del proceso terapéutico y posean un carácter flexible, promoviendo la integración y análisis de la información proveniente de diferentes profesionales, ámbitos y patologías. Esto es especialmente relevante en población infanto-juvenil debido no solamente sus crecientes dificultades y problemáticas<sup>16</sup> sino también a su heterogeneidad y comorbilidad (en torno al 50%).

Una de las mejores propuestas para solventar las limitaciones citadas se encuentra en el Sistema CORE, desarrollado en 1988 por el *CORE System Group* en el Reino Unido. Este sistema procura el cumplimiento de los requisitos expuestos anteriormente, permitiendo además demostrar la funcionalidad de las intervenciones dentro del contexto psicoterapéutico<sup>17</sup>. El CORE-OM es su instrumento principal, el cual ha experimentado una expansión extraordinaria desde su creación, existiendo versiones para más de 20 países, incluida la versión en español<sup>18</sup> que ha mostrado propiedades psicométricas satisfactorias<sup>19</sup>.

Finalmente, dentro de este innovador sistema se encuentra el *Young-Person's Clinical Outcomes in Routine Evaluation* (YP-CORE), la versión para niños y adolescentes. Para realizar el YP-CORE se partió del CORE-OM, el cual fue confeccionado tras revisar 638 ítems de los instrumentos más empleados en el Reino Unido como medida de resultado<sup>5</sup>. El CORE-OM cubre los siguientes aspectos, todos ellos relacionados con el bienestar psicológico: depresión, ansiedad, trauma, síntomas físicos, sensación subjetiva de malestar, dificultad en las relaciones y riesgo de agresión hacia uno mismo o a los demás. Los 34 ítems del CORE-OM fueron evaluados por una muestra de 48 profesionales y 45 jóvenes, dando lugar a una versión preliminar que después fue testada en una prueba piloto de 343 participantes, de la que finalmente surgió la versión final del YP-CORE<sup>7</sup>. El objetivo principal de este estudio es establecer las propiedades psicométricas de su versión en español, concluyendo así el proceso de adaptación. Con ello se pretende que el instrumento resulte utilizable para todos los servicios de atención psicológica y/o psiquiátrica infanto-juvenil de habla hispana que deseen evaluar el resultado de sus intervenciones, sin perjuicio de que sean necesarias futuras adaptaciones a contextos culturales y lingüísticos específicos.

La versión española del YP-CORE ya se encuentra disponible para su descarga en formato de libre acceso, se puede encontrar tanto en <http://www.ub.edu/terdep/core/> como en <https://www.coresystemtrust.org.uk/translations/spanish>.

## METODOLOGÍA

### Adaptación del instrumento

El estudio presentado, además de cumplir los requisitos estipulados por la *International Test Commission*<sup>20</sup>, ha tenido particularmente en cuenta las premisas especificadas por el CORE System Trust (CST)<sup>21</sup> para la adaptación de sus instrumentos a otras lenguas. Entre éstas destacan la importancia de involucrar la participación un grupo heterogéneo de personas y preservar el significado original de sus ítems.

Siguiendo estas directrices, se contactó con el responsable (C.E.) del *CORE System Group* para establecer los acuerdos iniciales y planificar el procedimiento. Seguidamente se inició el proceso de traducción del YP-CORE al español, en el que participaron una traductora profesional, un experto en salud mental y cuatro personas legas, todas ellas con un amplio conocimiento de la lengua inglesa. A continuación, se convocó una reunión con C.E., tres profesionales de salud mental y una persona lega de 14 años de edad, de la cual se tomaron muy en cuenta las aportaciones acerca de la comprensión de la prueba. Se discutieron las diferentes versiones procurando concretar cual representaba mejor a la original y era más aplicable para población infanto-juvenil hispanohablante.

Esta versión preliminar fue sometida a la valoración de 7 jóvenes hispanohablantes, todos ellos con un nivel socio-económico medio. Las distintas retroalimentaciones fueron enviadas a C.E. y se consideraron conjuntamente algunas pequeñas variaciones. De la versión resultante realizaron retrotraducciones dos personas bilingües, sin que se apreciaran diferencias relevantes respecto a la versión original.

### Participantes

La muestra no clínica la compusieron un total 136 participantes en la primera aplicación y 128 en la segunda, todos ellos alumnos de los cuatro cursos de enseñanza secundaria y del sexto curso de primaria. El cuestionario se les aplicó en dos momentos temporales distintos, con una diferencia de 14 días.

La muestra clínica se compuso de un total de 122 chicos y chicas. En esta muestra se encontraron registrados un amplio abanico de diagnósticos, agrupados a partir de los criterios diagnósticos de los centros y las dimensiones propuestas por el DSM-IV-TR<sup>22</sup>. En total se utilizaron 11 categorías diferentes, que aparecieron, combinadas o no, en 196 ocasiones. Cabe resaltar la elevada comorbilidad, ya que se pudieron registrar un mínimo de 2 trastornos simultáneos ( $n=35$ , 28,69%) y un máximo de hasta 4 ( $n=1$ , 0,82%).

Los criterios de inclusión fueron tener entre 11 y 16 años, un buen nivel de comprensión del español, no estar diagnosticado de un Trastorno Psicótico o Trastorno del Espectro Autista (TEA) y, específicamente en la muestra clínica, haber sido diagnosticado de alguno de los trastornos recogidos en el DSM-IV-TR.

De los 258 participantes iniciales, 10 fueron excluidos, todos ellos de la muestra clínica; seis por superar la franja de edad estipulada (uno de los cuales además estaba diagnosticado de Trastorno Psicótico) y 4 por presentar el diagnóstico de TEA.

### Instrumentos

- YP-CORE<sup>7</sup>: Instrumento breve en formato de autoinforme presentado como una medida genérica capaz de detectar el malestar generado por un amplio rango de problemas, a la vez que proporciona información sobre el funcionamiento general de la persona. La prueba está diseñada para ser contestada por personas de entre 11 y 16 años y se compone de un total de 10 ítems (un ítem de Riesgo, un ítem de Bienestar Subjetivo, cuatro ítems de Problemas/Síntomas y cuatro de Funcionamiento General), que deben responderse en una escala de formato gradual tipo Likert de 5 puntos (de 0 a 4). Por lo tanto, el rango de puntuaciones posibles va de 0 a 40, obteniéndose únicamente una puntuación total. Ésta se divide por el número de ítems que se han contestado para controlar los valores perdidos y facilitar la interpretación de los resultados. En 2015 se publicaron nuevos datos sobre sus propiedades psicométricas, los cuáles se han incluido en las comparaciones<sup>8</sup>.
- YSR<sup>12</sup>: Escala de autoinforme que permite la evaluación de competencias psicosociales (17 ítems), así como de problemas emocionales y comportamentales (112 ítems) en niños y adolescentes con edades comprendidas entre 11 y 18 años. Esta última parte contiene 14 ítems que describen conductas adaptativas o prosociales y 98 que se centran en conductas problemáticas. Según Achenbach<sup>23</sup>, la escala clasifica los trastornos psicológicos en dos síndromes de "banda ancha" (problemas de internalización y externalización) y ocho síndromes de "banda estrecha" (Aislamiento, Quejas somáticas, Ansiedad/Depresión, Problemas Sociales, Problemas de Pensamiento, Problemas de Atención, Conducta Delictiva y Comportamiento Agresivo). Los encuestados deben centrarse en los últimos seis meses, puntuando de 0 a 2. Finalmente, el grado de desajuste psicológico general se refleja en una puntuación total de conductas problema, y el grado de adaptación mediante una puntuación total de conductas deseables. El YSR se administró únicamente a la muestra clínica.

## Procedimiento

Una vez aprobado el estudio por el CST y los comités éticos correspondientes, el siguiente paso fue activar el proceso de administración de las pruebas en los distintos centros e instituciones. Para la muestra no clínica se obtuvo la aprobación del instituto *Institució Pedagògica Sant Isidor* (IPSI) de Barcelona, siendo la administración realizada por un estudiante de máster ante la presencia del profesor.

Para la muestra clínica se contactó con dos centros asociados a la *Universitat de Barcelona* (UB), el *Hospital de Mataró* y el *Institut de Trastorns Alimentaris* (ITA). En éstos la administración fue realizada por cuatro psicólogos clínicos infantiles y una psiquiatra antes o al terminar las sesiones terapéuticas grupales, informando y obteniendo el consentimiento siempre en primera instancia de padres y/o tutores legales. En todas las administraciones se expusieron los objetivos del estudio, la libertad para abstenerse y las características del cuestionario.

Una vez introducida la información en una base de datos, se consideró el uso de pruebas no paramétricas debido a la no normalidad de las puntuaciones<sup>24,25</sup>. Aun así, y reconociendo la potencia de los estadísticos paramétricos, éstos han sido utilizados adicionalmente en cada apartado para corroborar los resultados.

Para el análisis de los datos recogidos se ha utilizado el *Statistical Package for the Social Sciences* (SPSS) 21.0<sup>26</sup>. Paralelamente, para corregir las puntuaciones del test escogido con el objetivo de estudiar la validez convergente, el *Youth Self Report* (YSR), se ha utilizado el *Achenbach System of Empirically Based Assessment* (ASEBA). Además, para estudiar y realizar el Análisis de los Componentes Principales de la prueba se ha utilizado el programa *Monte Carlo PCA for Parallel Analysis*, posibilitando la realización del análisis paralelo.

## RESULTADOS

### Aceptabilidad

Un total de 235 de los participantes (94,8%) contestaron el instrumento completo. En el grupo no clínico el porcentaje fue del 96,3%, mientras que en la muestra clínica fue del 92,9%. Las diferencias entre ambos grupos no resultaron significativas ( $\chi^2_{(1, 248)}=1,486; p=0,223$ ). Concretamente, para la muestra no clínica hubo 3 (2,2%) cuestionarios con un ítem sin contestar, y dos (1,5%) con dos ítems en blanco. La media de ítems no contestados fue de 0,05. En cuanto a la muestra clínica, se devolvieron 6 cuestionarios (5,3%) con un ítem no contestado y un cuestionario (0,9%) con dos ítems no res-

Tabla 1		Edad y sexo de los participantes con la totalidad del YP-CORE contestado			
Muestra	n	Sexo (%)		Edad	
		Chicos	Chicas	Rango	Cuartiles
Clínica	104	30 (29%)	74 (71%)	11-16	14,15,16
No clínica	131	64 (49%)	67 (51%)	11-16	12,13,14
Total	235	94 (40%)	141 (60%)	11-16	13,14,15

pondidos. En tan solo uno (0,9) se observaron cinco ítems sin respuesta. En este caso la media de ítems no contestados fue de 0,12. El ítem menos contestado en ambas muestras fue el 5, con dos y tres omisiones. A partir de aquí, se han utilizado únicamente los datos de aquellas personas que completaron la totalidad del YP-CORE: un total de 235, 104 para la muestra clínica y 131 en la no clínica (véase Tabla 1).

### Pruebas de normalidad

A partir de los resultados obtenidos en asimetría y curtosis y en la prueba Kolmogorov-Smirnov (K-S), en la cual las puntuaciones fueron de 0,113 para la muestra no clínica y 0,094 para la muestra clínica ( $p<0,05$ ), puede afirmarse que las puntuaciones del YP-CORE siguieron una distribución alejada de la curva normal. Como ejemplo, puede observarse la distribución de la puntuación Total para la muestra no clínica en la Figura 1. En base a estos resultados se emplearon estadísticos no paramétricos como primera opción en los análisis.

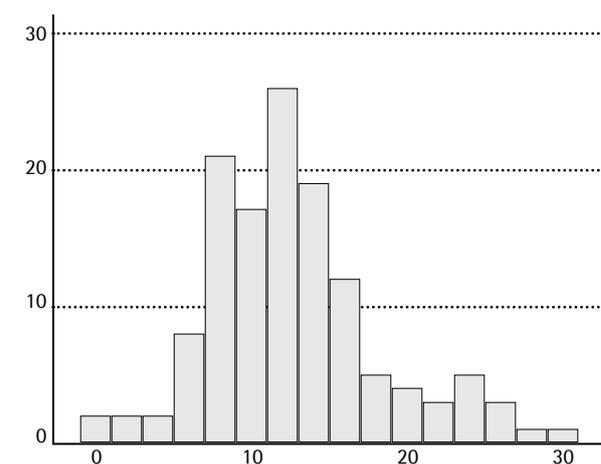


Figura 1 | Histograma de la puntuación Total en la muestra no clínica

## Fiabilidad

El instrumento demostró una buena consistencia interna, con valores alfa superiores a 0,7, tanto para la puntuación Total como Total sin Riesgo en el total de la muestra. Específicamente para los chicos de la muestra no clínica la fiabilidad para la puntuación Total fue 0,72 (IC 95%: 0,71–0,91), y para las chicas 0,85 (IC 95%: 0,79–0,89), resultados consistentes con los británicos<sup>8</sup>. En cuanto a la muestra clínica, los valores fueron 0,82 (IC 95%: 0,71–0,90) y 0,86 (IC 95%: 0,81–0,90) para chicos y chicas respectivamente.

## Estabilidad test-retest

De los 131 participantes de la muestra no clínica que lo completaron en su totalidad en la primera aplicación, 125 (95,42%) lo respondieron en la segunda. En este caso solo se tuvieron en cuenta los 120 (91,60%) que respondieron todos los ítems.

Siguiendo las líneas de Hinkle, Wiersma y Jurs<sup>27</sup>, se encontró una correlación moderadamente positiva entre la primera y la segunda aplicación del cuestionario, tanto para la puntuación Total ( $\rho$  de Spearman=0,638;  $r$  de Pearson=0,645) como Total sin Riesgo ( $\rho$ =0,637;  $r$ =0,644).

Seguidamente, el estadístico de Wilcoxon reflejó diferencias significativas entre las dos aplicaciones, tanto para la puntuación Total ( $Z=3,053$ ,  $p<0,05$ ) como para la puntuación Total menos Riesgo ( $Z=3,083$ ,  $p<0,05$ ), un efecto que suele observarse en estudios con aplicaciones repetidas de un mismo instrumento de evaluación<sup>28</sup>. Además, si atendemos a las puntuaciones medias, estas diferencias muestran una mejora en la sintomatología dentro de la población no clínica, aunque tanto la correlación de Pearson ( $r=0,20$ ) como la  $d$  de Cohen ( $d=0,26$  y  $0,27$ ) indican que se trata de un tamaño del efecto pequeño para ambas escalas.

En cuanto a los estadísticos paramétricos, la prueba T de Student reflejó diferencias estadísticamente significativas entre la primera aplicación de la prueba y la segunda ( $t_{(119)}=2,847$ ,  $p<0,05$ ) atendiendo a la puntuación Total. Además, eliminando el ítem de Riesgo aparecieron resultados similares, existiendo diferencias estadísticamente significativas entre la primera aplicación y la segunda ( $t_{(119)}=2,938$ ,  $p<0,05$ ).

## Validez Convergente

Para llevar a cabo este punto del estudio se tuvieron en cuenta únicamente los 112 ítems pertenecientes a la parte de problemas emocionales y comportamentales del YSR, descartando aquellos cuestionarios con más de 11 ítems

perdidos al considerar que resultaría poco fiable para la interpretación de los análisis subsiguientes<sup>12</sup>. Siguiendo este criterio se tuvieron en cuenta un total de 95 participantes de la muestra clínica.

Inicialmente se analizó la consistencia interna de todas las dimensiones de la prueba, la cual resultó ser buena en la mayoría de escalas ( $0,8<\alpha<0,9$ ). Posteriormente se llevó a cabo el análisis de correlaciones entre ambos tests, interrelacionando todas las puntuaciones posibles del YSR con la puntuación Total y Total sin Riesgo del YP-CORE, encontrando correlaciones de entre 0,39 y 0,75 según el estadístico de Spearman ( $Rho$ ) y entre 0,36 y 0,75 con el de Pearson ( $r$ ), resultando todas significativas ( $p<0,001$ ). Concretamente, las correlaciones más elevadas se dieron entre las puntuaciones totales de ambos test, siendo 0,75 según ambos estadísticos, y entre la puntuación Total del YP-CORE y los problemas internalizantes de banda ancha, así como los afectivos y aislamiento/depresión de banda estrecha, con puntuaciones de 0,69, 0,70 y 0,70 respectivamente según el estadístico de Spearman, y 0,72, 0,72 y 0,71 según el de Pearson.

## Influencia del sexo

Se observaron diferencias estadísticamente significativas, tanto en la muestra no clínica ( $Z=-2,08$ ,  $p=0,037$ , para la puntuación Total;  $Z=-1,99$ ,  $p=0,047$ , para Total sin riesgo) como en la clínica ( $Z=-2,33$ ,  $p=0,020$ , para la puntuación Total;  $Z=-2,21$ ,  $p=0,027$ , para Total sin riesgo), con puntuaciones significativamente más elevadas en las chicas. Los tamaños del efecto para esta diferencia son más elevados en la muestra clínica ( $d=0,52$  y  $r=0,23$ , para la puntuación Total;  $d=0,48$  y  $r=0,22$ , para Total sin riesgo) que en la no clínica ( $d=0,35$  y  $r=0,18$ , para la puntuación Total;  $d=0,34$  y  $r=0,17$  para Total sin riesgo). En este punto se observó además la única diferencia entre estadísticos en la puntuación Total menos Riesgo para la muestra no clínica, donde la prueba T para muestras independientes ( $t_{(129)}=-1,94$ ,  $p=0,055$ ) no corroboró los resultados de la prueba U de Mann-Whitney.

## Influencia de la edad

Los resultados mostraron la no existencia de correlaciones significativas tanto en la muestra clínica ( $Rho=0,16$  y  $r=0,19$ ) como en la no clínica ( $Rho=0,09$ ;  $r=0,10$  y  $0,11$ ), para las puntuaciones totales y sin el ítem de riesgo respectivamente. A diferencia de los resultados británicos<sup>8</sup>, aunque en nuestros participantes se ha encontrado un ligero aumento de las puntuaciones en función de la edad en la muestra clínica, este no es estadísticamente significativo. Estudiando

ambos sexos por separado las correlaciones tampoco fueron significativas en ninguna de las dos muestras.

### Diferencias entre la muestra clínica y no clínica

Los resultados de la prueba U de Mann-Whitney reflejan diferencias estadísticamente significativas entre ambas muestras tanto para la Puntuación Total como Total sin Riesgo, mostrando una mayor puntuación en la muestra clínica. Aun así, existe discrepancia entre los tamaños del efecto calculados (véase Tabla 2).

La prueba T de Student para muestras independientes ha corroborado estas diferencias. De esta forma, los participantes clínicos obtienen una puntuación significativamente mayor en comparación con la muestra no clínica ( $t_{(233)} = -5,95$ , IC 95% = -7,53 - -3,78,  $p < 0,001$ ) atendiendo a la puntuación total del test. Además, sin el ítem de Riesgo se muestran igualmente diferencias estadísticamente significativas ( $t_{(233)} = -5,55$ , IC 95% = -7,21 - -3,43,  $p < 0,05$ ).

A partir de los resultados anteriores, se consideró relevante comparar las puntuaciones por sexo. Mediante la prueba U de Mann-Whitney, aparecieron diferencias estadísticamente significativas entre los chicos de ambas muestras ( $Z = -2,26$ ,  $p = 0,024$  para la puntuación Total;  $Z = -2,06$ ,  $p = 0,039$  para Total sin riesgo) y entre las chicas ( $Z = -4,33$  y  $Z = -4,08$  para la puntuación Total y para el Total sin riesgo respectivamente,  $p < 0,0005$ ).

Por otro lado, la prueba T de Student para muestras independientes reflejó igualmente diferencias estadísticamente significativas entre chicos ( $t_{(92)} = -3,68$ , IC 95% = -6,78 - -0,58,  $p < 0,05$  para la puntuación Total;  $t_{(92)} = -3,58$ , IC 95% = -6,82 - -0,35,  $p < 0,05$  para el Total sin riesgo) y entre chicas ( $t_{(139)} = -5,91$ , IC 95% = -8,26 - -3,55,  $p < 0,05$  para la puntuación Total;  $t_{(139)} = -5,46$ , IC 95% = -7,82 - -3,10,  $p < 0,05$  para el Total sin riesgo). Al igual que en la comparación entre la totalidad de las muestras, y como puede observarse en la Tabla 2, aparecieron discrepancias entre los tamaños del efecto calculados, siendo aun así más elevados para la muestra femenina.

### Análisis de los Componentes Principales

Atendiendo a los resultados reflejados por el análisis paralelo, pueden retenerse dos factores para ambas muestras. El uso del análisis paralelo se debe al soporte empírico que acumula, en contraposición a las críticas originadas por otros métodos<sup>29</sup>. Así, tanto para la muestra clínica como para la no clínica los ítems que representan cada factor son los mismos, variando únicamente en el resultado de su valor propio.

Al igual que en el estudio original<sup>7</sup>, las agrupaciones resultantes de la rotación oblicua (véase Tabla 3) sugieren un componente formado por los ítems redactados de forma positiva (ítems 3, 5 y 10) y otro por los ítems redactados de forma negativa (ítems 1, 2, 4, 6, 7, 8 y 9). Además, cabe resaltar que la correlación entre factores para la muestra clínica fue casi idéntica a la del estudio original (0,36).

Tabla 3		ACP para ambas muestras			
		Muestra no clínica		Muestra clínica	
Ítems	Factores	Ítems	Factores		
	1	2	1	2	
7	0,752	1	0,885		
6	0,701	6	0,847		
9	0,685	9	0,787		
2	0,647	7	0,718		
1	0,616	2	0,643		
8	0,500	8	0,570		
4	0,493	4	0,505		
3		0,721	10	0,838	
5		0,666	3	0,708	
10		0,624	5	0,673	

Tabla 2		Diferencia de puntuaciones entre la muestra clínica y no clínica							
Medida	No clínica		Clínica		Diferencia Media	Z (U Mann-Whitney)	p	r	d Cohen
	M	DT	M	DT					
Total	12,18	5,55	17,84	5,96	-5,65 (-7,53 - -3,78)*	-5,31	0,000	0,35	0,98
Total no R	13,37	8,34	18,69	8,21	-5,32 (-7,21 - -3,43)*	-5,02	0,000	0,33	0,64

\* Intervalos de Confianza del 95%.

Tabla 4	Puntos de Corte para la puntuación Total y Total menos Riesgo	
	Grupo	PC Total
Chicos	12,72	13,88
Chicas	15,49	16,64
Muestra Total	14,91	16,05

### Puntos de corte y cambio fiable y clínicamente significativo

Al existir diferencias estadísticamente significativas entre sexos, los puntos de corte se han calculado para ambos grupos por separado. Además, también se proporcionan eliminando el ítem de riesgo, donde se refleja un ligero aumento de los puntos proporcionados (véase Tabla 4). En relación al estudio británico de 2015, los resultados replican las diferencias encontradas entre sexos, con puntos de corte más elevados para las chicas (casi 3 puntos por encima). Los puntos de corte de la Puntuación Total menos Riesgo se sitúan ligeramente por encima, pudiéndose interpretar que los participantes han tendido a puntuar bajo en el ítem de Riesgo.

Del mismo modo, la estimación del cambio fiable también se ha realizado para ambos sexos por separado, utilizando el total de la muestra. En el caso de los chicos, el valor del cambio fiable se situó en 8,17 para la puntuación Total y en 8,83 para la Total sin riesgo, mientras que para las chicas fue de 8,86 en el caso de la puntuación Total y de 9,46 en el caso de la Total sin riesgo. Por lo tanto, cualquier cambio superior a los indicados en cada caso se puede considerar un cambio fiable. En el artículo británico se dan los valores tan solo para la puntuación Total, siendo 8,2 para los chicos y 7,7 para las chicas. Con respecto a estos resultados, se observa una gran similitud en el caso de los chicos, aunque en el caso de las chicas la diferencia es de más de un punto.

### CONCLUSIONES

Este trabajo expone un primer estudio para la validación psicométrica al español del YP-CORE, un instrumento que resulta ser la medida más ampliamente utilizada en los servicios de asesoramiento psicológico (*counselling*) del Reino Unido<sup>30</sup>. Asimismo, en un metaanálisis se indica que es la medida más fiable a la hora de calcular el cambio producido por la terapia<sup>31</sup>.

La versión española del instrumento refleja unas buenas propiedades psicométricas, lo cual se traduce, en primer lugar, en un elevado porcentaje de respuesta, algo que puede atribuirse, entre otras cosas, a su brevedad. Además, al igual que

en el estudio británico del 2015, su consistencia interna ha sido buena tanto para la totalidad de las muestras como por sexos. Presenta una aceptable estabilidad temporal, la cual sin embargo ha mostrado una disminución de las puntuaciones estadísticamente significativas, aunque con tamaños del efecto bajos, algo que ya se observó por ejemplo en la validación del CORE-OM original<sup>32</sup>. El estudio de la validez convergente ha ofrecido correlaciones elevadas prácticamente entre todas las puntuaciones consideradas, aunque principalmente entre las totales de ambos cuestionarios, y con la de problemas internalizantes del YSR; dato que podría estar indicando la mayor adecuación del YP-CORE para medir aspectos relacionados con el malestar interno del individuo, más que con la dimensión de externalización. También parece registrar mayores puntuaciones en el sexo femenino, aunque la variabilidad en cuanto a la influencia de la edad entre el presente estudio y el británico<sup>8</sup> deja entrever la importancia de atender a las características demográficas en general, juntamente con el tamaño de muestra considerado. Además, estos análisis pueden estar reflejando cierta influencia del nivel socioeconómico, cultural o del desarrollo evolutivo<sup>33</sup>. La estructura factorial inicial ha quedado replicada por el ACP, reproduciéndose de manera casi exacta la distribución de ítems por factor sugerida en la versión original. Por otra parte, los puntos de corte han reflejado similitudes con los resultados británicos recientes, aunque han resultado más elevados entre los participantes españoles. Aquí debería considerarse la mediación de las características sociodemográficas de las muestras, aunque también podrían entrar en juego el efecto de la medicación o los diferentes diagnósticos, así como su comorbilidad y la gravedad de sus síntomas. En cuanto a la estimación del cambio fiable, los valores identificados en la muestra española coinciden con los británicos en el caso de los chicos, aunque no en el caso de las chicas. Finalmente, aunque el ítem de Riesgo es el que menor carga factorial tiene en ambas muestras, parece que su exclusión de la puntuación Total no conduce a resultados distintos en ninguno de los análisis, por lo que el uso de la puntuación Total menos Riesgo no parece ser relevante a nivel estadístico.

A pesar de las evidentes ventajas del YP-CORE, entre las limitaciones del presente estudio están un tamaño de muestras (tanto clínica como no clínica) moderadamente bajo, lo cual dificulta su generalización, así como el análisis de las puntuaciones en relación a aspectos clínicos relevantes como el diagnóstico o la medicación. Además, la comparación entre ambas muestras refleja cierto solapamiento. En este sentido, aquellos participantes no clínicos con puntuaciones mínimamente elevadas podrían entrar dentro del espectro clínico. Todo ello puede afectar los resultados sobre la capacidad discriminativa de la prueba. Además, no contar con datos posteriores a una determinada terapia constituye otra limitación, ya que permitiría valorar la sensibilidad al cambio del instrumento en un contexto clínico, más allá de estimar el cambio fiable mediante cálculos estadísticos.

En conclusión, el YP-CORE se presenta como una prometedora prueba a utilizar en un amplio abanico de servicios de atención psicológica infanto-juvenil. Los diferentes análisis han reafirmado su potencialidad de cara a detectar el malestar psicológico en esta población, además de permitir distinguir aquéllos con mayor gravedad. Su brevedad y fácil comprensión la hacen ideal para su uso rutinario de cara observar el progreso de la persona y el cambio producido por la intervención terapéutica. Aun así, se reconoce la necesidad de realizar estudios con muestras mayores que evalúen el efecto de las intervenciones terapéuticas con esta herramienta.

#### AGRADECIMIENTOS

Muy especialmente, se agradece la participación voluntaria y desinteresada de Alejandro Botella en el proceso de adaptación al español de la prueba. También se agradece la colaboración de la *Institució Pedagògica Sant Isidor* (IPSI) el *Hospital de Mataró* y el *Institut de Trastorns Alimentaris* (ITA). Este estudio ha sido financiado parcialmente por la *Generalitat de Catalunya* (ref. SGR717).

#### CONFLICTO DE INTERESES

Los autores no presentan conflicto de intereses.

#### BIBLIOGRAFÍA

- Collishaw S, Maughan B, Goodman R, Pickles A. Time trends in adolescent mental health. *J Child Psychol Psychiatr.* 2004; 45(8):1350-62.
- Lemos S. La psicopatología de la infancia y la adolescencia: consideraciones básicas para su estudio. *Pap Psicol.* 2003; 24(85):19-28. Disponible en: [http://www.redalyc.org/pdf/778/77808503.pdf?origin=publication\\_detail](http://www.redalyc.org/pdf/778/77808503.pdf?origin=publication_detail)
- Caballo VE, Simón MÁ. Manual de psicología clínica infantil y del adolescente: trastornos específicos. Madrid: Ediciones Pirámide; 2002.
- Bados A, García E, Fusté A. Eficacia y utilidad clínica de la terapia psicológica. *Int J Clin Health Psychol.* 2002;2(3):477-502.
- Evans C, Mellor-Clark J, Margison F, Barkham M, Audin K, Connell J, et al. CORE: Clinical outcomes in routine evaluation. *J Mental Health.* 2000;9:247-55.
- del Barrio V. Evaluación psicológica en la infancia y la adolescencia. Madrid: Universidad Nacional de Educación a Distancia; 2003.
- Twigg E, Barkham M, Bewick BM, Mulhern B, Connell J, Cooper M. The Young Person's CORE: Development of a brief outcome measure for young people. *Couns Psychother Res.* 2009;9(3):160-8.
- Twigg E, Cooper M, Evans C, Freire E, Mellor-Clark J, McInnes B, et al. Acceptability, reliability, referential distributions and sensitivity to change in the Young Person's Clinical Outcomes in Routine Evaluation (YP-CORE) outcome measure: replication and refinement. *Child Adolesc Ment Health.* 2015;21(2):115-23.
- Kovacs M. Children's depression inventory (CDI2). North Tonawanda, NY: Multi-Health Systems, Inc.; 2011.
- Spielberger CD, Edwards CD, Montuori J, Lushene R. State-Trait Anxiety Inventory for Children. Palo Alto, CA: Consulting Psychologist Press; 1973.
- Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatr.* 1997;38(5):581-6.
- Achenbach T. Manual for the youth self-report and 1991 profile. Burlington, VT: University of Vermont Department of Psychiatry; 1997.
- Achenbach T, Edelbrock C. Child Behavior Checklist: CBCL. Burlington, VT: University of Vermont Department of Psychiatry; 1980.
- Achenbach T, Rescorla L. ASEBA school-age forms & profiles. Burlington, VT: ASEBA; 2001.
- Barkham M, Evans C, Margison F, McGrath G, Mellor-Clark J, Milne D, et al. The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *J Ment Health.* 1998;7(1):35-47.
- Forns M. Evaluación psicológica infantil. Barcelona: Barcanova; 1993.
- Barkham M, Mellor-Clark J, Connell J, Cahill J. A core approach to practice-based evidence: A brief history of the origins and applications of the CORE-OM and CORE System. *Couns Psychother Res.* 2006;6(1):3-15.
- Feixas G, Evans C, Trujillo A, Saúl LA, Botella L, Corbella S, et al. La versión española del CORE-OM: Clinical Outcomes in Routine Evaluation-Outcome Measure. *Rev Psicoter.* 2012;89:1-22.
- Trujillo A, Feixas G, Bados A, Garcia-Grau E, Salla M, Medina JC, et al. Psychometric properties of the spanish version of the clinical outcomes in routine evaluation-Outcome Measure. *Neuropsychiatr Dis Treat.* 2016;12:1457-66.
- Hambleton RK, Merenda PF, Spielberger CD. Adapting educational and psychological tests for cross-cultural assessment. Tilburg: Psychology Press; 2005.
- Evans C. Translating and "normalising" CORE system: CORE System Trust (CST) position statement; 2008. Disponible en: [http://www.psych.org/CORE-OM/Translating\\_CORE-OM\\_CST\\_position\\_statement.pdf](http://www.psych.org/CORE-OM/Translating_CORE-OM_CST_position_statement.pdf)
- Michael B, Harold PA, Allen F. DSM-IV TR. Manual diagnóstico y estadístico de los trastornos mentales. Barcelona: Masson; 2002.
- Achenbach T. Manual for the youth self-report and 1991 profile. Burlington, VT: University of Vermont Department of Psychiatry; 1991.
- Gómez-Gómez M, Danglot-Banck C, Vega-Franco L. Sinopsis de pruebas estadísticas no paramétricas. Cuando usarlas. *Rev Mex Pediatr.* 2003;70(2):91-9.
- Guàrdia J, Però M, Freixa M, Turbany J. Análisis de datos en psicología (2ª ed.). Barcelona: Delta Publicaciones; 2008.
- IBM Corporation. IBM SPSS statistics for Windows, Version 21.0. Armonk, NY: IBM Corp; 2012.
- Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. Michigan: Houghton Mifflin; 2003.
- Durham C, McGrath L, Burlingame G, Schaalje G, Lambert M, Davies D. The effects of repeated administrations on self-report and parent-report scales. *J Psychoeduc Assess.* 2002; 20(3):240-57.
- Horn J. A rationale and test for the number of factors in factor analysis. *Psychometrika.* 1965;30(2):179-85.
- Cooper M, Pybis J, Hill A, Jones S, Cromarty K. Therapeutic outcomes in the Welsh Government's school-based counselling strategy: An evaluation. *Couns Psychother Res.* 2013;13(2):86-97.
- Cooper M. Counselling in UK secondary schools: A comprehensive review of audit and evaluation data. *Couns Psychother Res.* 2009;9(3):137-50.
- Evans C, Connell J, Barkham M, Margison F, McGrath G, Mellor-Clark J, et al. Toward a standardised brief outcome measure: Psychometric properties and utility of CORE-OM. *Br J Psychiatry.* 2002;180(1):51-60.
- Greig A, Taylor J, MacKay T. Doing research with children: A practical guide. Glasgow: Sage; 2012.