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

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

[1]Department of Clinical Psychology and Psychobiology, Universitat de Barcelona
[2]Institut de Neurociències, Universitat de Barcelona
[3]Institut de Trastorns Alimentaris, Barcelona
[4]Blanquerna Faculty of Psychology, Education and Sport Sciences, Universitat Ramon Llull, Barcelona
[5]Psychology Department, University of Roehampton, London

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

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

Correspondence:
Guillem Feixas
Departamento de Psicología Clínica y Psicobiología, Facultad de Psicología,
Universitat de Barcelona
Passeig Vall d'Hebron, 171, 08035 Barcelona, Spain
Tel.: (+34) 933125100
E-mail: gfeixas@ub.edu

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

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

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

INTRODUCTION

There is empirical evidence on the growing prevalence of mental health problems among children and adolescents[1], estimated to be around 17-20%[2]. In order to address this situation, strategies to improve mental health in these populations are particularly relevant, especially treatment programs suitable for the wide range of problems that young people may experience[3]. With growing need, it is vital to use demonstrable effective and clinically useful resources[4], to prevent increasing biopsychosocial impairment and to improve teenagers' mental health and quality of life.

In this light it is important to monitor change throughout the therapeutic process to give more extensive evidence of progress and to allow comparisons between approaches, centres, and practitioners[5]. Existing measures of change are, for instance, direct observation, interviews, narrative self-report and, above all, projective techniques[6].

Nevertheless, these techniques may be insufficient and are best complemented with assessment instruments allowing quantitative record of changes occurred and potentially supporting the detection of any necessary modification in the therapeutic intervention. Many of the instruments developed so far are focused on measuring the symptomatology of each pathology in particular, being poorly applicable to patients with other problems and particularly unsatisfactory with the unclear or uncertain diagnoses[7], which are fairly frequent in the adolescence. As a result, there are few measures validated for children and adolescents, with most tailored for high-severity cases[8]. Among the instruments most used nowadays, there are the Children Depression Inventory[9] (CDI), the State-Trait Anxiety Inventory for Children[10] (STAIC), the Strengths and Difficulties Questionaire[11] (SDQ), and the Youth Self Report[12] (YSR), derived from the Child Behaviour Checklist[13] (CBCL), among others. The Spanish version of the YSR, authored by Achenbach and Rescorla[14], was used in this study.

The situation reported above describes the need for brief, general instruments, which avoid the use of multiple tools and resources[15], foster the quantitative assessment of the therapeutic process and are flexible in nature. Measures with these features promote integration and analysis of information stemming from different professionals, areas, and pathologies, and are especially relevant in young people, not only because of their increasing difficulties and problems[16], but also due to heterogeneity of problems and prevalent comorbidity (around 50%).

One of the best proposals to solve these issues is the CORE (Clinical Outcomes in Routine Evaluation) System, developed in 1988 by the CORE System Group in the United Kingdom. This system tries to fulfil all the requirements described in the previous paragraphs and supporting assessment of the effectiveness of interventions within psychotherapeutic contexts[17]. The CORE-OM is the main instrument of this system, which has experienced an extraordinary expansion since its creation. There are versions for more than 20 countries, including Spain[18], where it has shown satisfactory psychometric properties[19].

The CORE system now includes the Young-Person's equivalent of the CORE, the YP-CORE, designed for children and adolescents. The YP-CORE derives from the CORE-OM, which itself was designed after reviewing 638 items from the outcome measures most widely used instruments in the United Kingdom[5]. The CORE-OM covers the following areas, all related to psychological distress: depression, anxiety, trauma, physical symptoms, subjective feeling of distress, difficulties in relationships, and risk of aggression towards oneself and others. The 34 items of the CORE-OM were revised by a sample of 48 professionals and 45 youngsters, leading to a preliminary version which was tested in a pilot study with 343 participants, a process after which the final version of the YP-CORE was finally obtained[7]. The main objective of this study is to establish the psychometric properties of its Spanish version, concluding the adaptation process. The rationale for this procedure is to make the instrument usable for all Spanish-speaking psychological or psychiatric healthcare services for children and adolescents willing to assess the result of their interventions, without prejudice of further necessary cultural or linguistic adaptations to specific contexts.

The Spanish version of the YP-CORE is already available to download in open-access. It may be found both at http://www.ub.edu/terdep/core/ and https://www.coresystemtrust.org.uk/translations/spanish.

METHODS

Adaptation of the instrument

The study met all the requirements of the International Test Comission[20] and the criteria specified by the CORE System Trust (CST)[21] for translation of their instruments. These standards emphasise the importance of involving a heterogeneous group of people in the adaptation process, and also the need to preserve the original meaning of items.

Following these guidelines, the CORE System Group representative (C.E.) was contacted to establish the initial agreement and plan the procedure. Afterwards, the translation process of the instrument into Spanish began, asking one professional translator, one mental health expert, and four lay people to make their own translations. All these collaborators were proficient English-speakers. Following this step, a meeting was arranged with C.E., three mental health professionals, and a lay person aged 14, whose opinions about the understanding of the questionnaire were especially taken into account. All the translations were considered, discussing about which one attained best the original meaning of every item, and was more applicable to the Spanish-speaking population.

The meeting led to a preliminary version of the test, which was further reviewed by 7 Spanish-speaking youths from a medium socioeconomic background. Feedback was collected from them and discussed with C.E., producing slight changes in some items. Finally, two different bilingual persons back-translated the questionnaire into English, a final step that did not reveal any major differences from the original version.

Participants

The non-clinical sample comprised 136 participants in the first administration of the test and 128 in the second. All were students from the sixth year of primary school or the four years of secondary school. The questionnaire was applied twice, with an interval of 14 days.

The clinical sample was composed of 122 participants with a wide range of diagnoses. These were grouped according to the diagnostic criteria provided by their healthcare centre and dimensions proposed by the DSM-IV-TR[22]. In total, 11 different diagnostic categories were used, which were registered, either combined or standalone, in 196 occasions. It is important to underline the high comorbidity, since the number of simultaneous disorders ranged from 2 (n=35, 28.69%) to 4 (n=1, 0.82%).

Inclusion criteria were: being between 11 and 16 years old, good level of Spanish understanding, not being diagnosed with any psychotic or autism spectrum disorder (ASD) and, for the clinical sample, having been given at least one other diagnosis in the DSM-IV-TR except the aforementioned.

From the 258 initial participants, 10 were excluded, all of them from the clinical sample; 6 for not meeting the age criterion (one of them was additionally diagnosed with a psychotic disorder), and 4 having been diagnosed with an ASD.

Instruments

- YP-CORE[7]: This is a brief self-report instrument used as a generic measure to detect distress generated by a wide range of problems and to provide information about the person's general functioning. The questionnaire is designed to be filled by people between 11 and 16 years old. It is composed of 10 items (1 Risk item, 1 Wellbeing item, 4 Problems/Symptoms items, and 4 General Functioning items), which are answered with a 5-point Likert scale (from 0 to 4). Thus, the possible scores range from 0 to 40, obtaining an overall score. This score is divided by the number of completed items in order to control missing values and facilitate interpretation of the results. In 2015, the British YP-CORE psychometric properties were updated with new published data, which have been used for the comparisons in this article[8].

- YSR[12]: Self-report scale to assess psychosocial competencies (17 items), and emotional and behavioural problems (112 items) among children and adolescents aged between 11 and 18 years. This second part includes 14 items describing adaptive or prosocial behaviours, and 98 referring to problematic behaviours. According to Achenbach[23], the scale classifies disorders into two "broadband" syndromes (internalising and externalising problems), and eight "narrowband" syndromes (withdrawn, somatic complaints, anxiety/depression, social problems, thought problems, attention problems, rule-breaking behaviours, and aggressive behaviours). Respondents are asked to focus on the last six months, scoring from 0 to 2. Finally, the overall degree of psychological imbalance is summarised in a total score of problematic behaviours, while the degree of adaptation is summarised in a total score of desirable behaviours. In this study, the YSR was only administered to the clinical sample.

Procedure

Once the study was approved by the CST and the ethics committees of the organisations involved, the next step was to start the administration of questionnaires at the collaborating centres. For the non-clinical sample, the school *Institució Pedagògica Sant Isidor* (IPSI), from Barcelona, agreed to participate. Data collection was conducted by a masters' student with the teacher present at the classroom.

For the clinical sample, two healthcare centres, partners of the *Universitat de Barcelona* (UB), were contacted: *Hospital de Mataró* and *Institut de Trastorns Alimentaris* (ITA). At these centres, the administration was conducted by four child clinical psychologists and one psychiatrist at the beginning or the end of therapeutic group sessions, always informing and obtaining the consent of all participants' par-

ents or tutors first. Before any answer was collected, the study aims, information about their freedom to choose whether to participate or not, and the questionnaire characteristics were highlighted.

Non-parametric statistics were considered to be the best option due to the non-normality of the questionnaire's scores[24,25]. But parametric statistics are also reported given their greater statistical power and to support comparability between these and other results.

For the analysis of the data collected, the Statistical Package for the Social Sciences (SPSS) 21.0[26] was used. Additionally, to compute the YSR scores from the test applied to estimate the convergent validity, the, the Achenbach System of Empirically Based Assessment (ASEBA) was used. Finally, for Principal Component Analysis (PCA) of the questionnaire, the software Monte Carlo PCA for Parallel Analysis was also used to compute a parallel analysis.

RESULTS

Acceptability

A total of 235 participants (94.8%) completed the entire questionnaire. In the non-clinical group, the number was 131 (96.3%), whereas in the clinical one the number was 104 (92.9%). The difference between samples in non-completion was not significant ($\chi^2_{(1, 248)}$=1.486; p=0.223). In the non-clinical sample there were 3 (2.2%) questionnaires with one missing item, and 2 (1.5%) with two missing items. The mean percentage of missing items was 0.05%. With regard to the clinical sample, 6 questionnaires (5.3%) were returned with one missing item, 1 questionnaire (0.9%) with two missing items and one participant (0.9%) omitted five items, giving an overall percentage of missing items of 0.12%. The most common missing item in both samples was the 5th, with two and three omissions respectively. For the rest of the analyses, data from only the completers of the entire questionnaire have been used (see Table 1).

Normality

Considering the skewness and kurtosis results and the Kolmogorov-Smirnov test (*K-S*) for which the scores were 0.113 for the non-clinical sample and 0.094 for the clinical (p<0.05) it can be concluded that the scores of the YP-CORE do not follow a Gaussian distribution. As an example, the distribution of the Total score in the non-clinical sample may be seen in Figure 1. As noted above, for this reason, non-parametric statistics were the first choice for the analyses.

Table 1 Age and sex of the participants who completed the whole YP-CORE

Sample	n	Sex (%)		Age	
		Boys	Girls	Range	Quartile
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

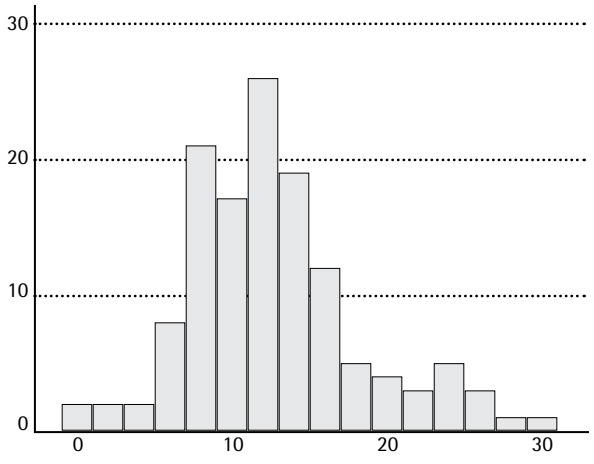

Figure 1 *Histogram of the Total score in the non-clinical sample*

Reliability

The YP-CORE showed good internal consistency, with Cronbach's Alpha values higher than 0.7, both for the Total and for the Total except Risk scores in the whole sample. Specifically, for the boys in the non-clinical sample, the reliability of the Total score was 0.72 (IC 95%: 0.71–0.91), while for the girls it was 0.85 (IC 95%: 0.79–0.89). These results are consistent with the British[8]. For the clinical sample is concerned, the values were 0.82 (IC 95%: 0.71–0.90) and 0.86 (IC 95%: 0.81–0.90) for boys and girls respectively.

Test-retest stability

Among the 131 participants in the non-clinical sample who completed the entire questionnaire at the first administration, 125 (95.42%) filled it in the second time. Data

from the 120 (91.60%) participants who completed the whole questionnaire are reported here.

Following Hinkle, Wiersma, and Jurs[27] categories, a moderately positive correlation was found between the first and second administrations of the questionnaire, both for the Total score (Spearman's *rho*=0.638; Pearson's *r*=0.645) and for the Total except Risk score (*rho*=0.637; *r*=0.644).

The Wilcoxon test statistic showed statistically significant differences between the two administrations, both for the Total score (Z=3.053, p<0.05) and for the Total except Risk score (Z=3.083, p<0.05), an effect that has previously been found in other studies with repeated administrations of the same assessment measure[28]. Furthermore, if we consider the mean scores, these differences show an improvement of the symptomatology within the non-clinical sample, although both Pearson correlation (r=0.20 for both scores) and Cohen's *d* (d=0.26 and 0.27 respectively) suggest a small effect size for both scores.

Regarding the use of parametric statistics, the Student's t test also showed statistically significant differences in the Total score between the first and the second administration of the test ($t_{(119)}$=2.847, p<0.05). In the case of the Total except Risk score the results were similar, founding statistically significant differences between the first and second administrations ($t_{(119)}$=2.938, p<0.05).

Convergent validity

In order to conduct this analysis, only the 112 items belonging to the second part of the YSR, assessing emotional and behavioral problems, were taken into account. We discarded those questionnaires with more than 11 missing items, since any interpretation of subsequent analyses would be unreliable[12]. Following this criterion, 95 participants of the clinical sample were included.

Initially, the internal consistency of all dimensions of the test was analysed, which proved to be good for most scales ($0.8 < \alpha < 0.9$). Subsequently, the correlation analysis between the two tests was performed interrelating all the possible YSR scores with the YP-CORE Total and Total except Risk scores. According to the Spearman statistic (*rho*), we found correlations between 0.39 and 0.75, while with Pearson statistic (r), the values were from 0.36 to 0.75, all highly statistically significant (p<0.001). The highest correlations were found between the Total scores of both tests (0.75 in both statistics), and between the YP-CORE Total score and the Internalizing broadband problems of the YSR, as well as the Affective and Isolation/Depression narrowbands, with scores of 0.69, 0.70, and 0.70 respectively according to the Spearman statistic, and 0.72, 0.72, and 0.71 according to Pearson.

Influence of gender

Statistically significant differences between males and females were observed in both the non-clinical (Z=-2.08, p=0.037, for the Total score; Z=-1.99, p=0.047, for the Total except Risk score) and the clinical samples (Z=-2.33, p=0.020, for the Total score; Z=-2.21, p=0.027, for the Total except Risk score), with higher scores in girls. The effect sizes for this difference were larger in the clinical sample (d=0.52 and r=0.23 for the Total score; d=0.48 and r=0.22 for the Total except Risk score) than in the non-clinical sample (d=0.35 and r=0.18 for the Total score; d=0.34 and r=0.17 for the Total except Risk score). This was the only test for which parametric and non-parametric statistics gave different results: for the Total except Risk score in the non-clinical sample the t-test was not statistically significant ($t_{(129)}$=-1.94, p=0.55) where the Mann-Whitney U test had been.

Influence of age

The results showed the absence of significant correlations in both the clinical (*rho*=0.16 and r=0.19) and the non-clinical samples (*rho*=0.09; r=0.10 and 0.11), in Total scores and Total except Risk scores respectively. Although in our participants a slight age-related increase in scores was found in the clinical sample, this was not statistically significant, unlike British results[8]. Analysing age separately within gender, the correlations were also not significant in any of the two samples.

Differences between clinical and non-clinical samples

The results of the Mann-Whitney U test showed statistically significant differences between the two samples in both Total and Total except Risk scales, with higher scores in the clinical sample. Nevertheless, there are discrepancies between the effect sizes (see Table 2).

The Student's t corroborated those differences. Taking into account the Total score of the test, clinical participants obtained significantly higher scores compared to the non-clinical sample ($t_{(233)}$=-5.95, IC 95%=-7.53 to -3.78, p<0.001). In addition, without the Risk, item there were also statistically significant differences ($t_{(233)}$=-5.55, IC 95%= -7.21 to -3.43, p<0.05).

In the light of the effect of gender noted above, it was considered relevant to explore the clinical *vs.* non-clinical difference within gender. The Mann-Whitney U test showed statistically significant differences within the boys (Z=-2.26, p=0.024 for the Total score; Z=-2.06, p=0.039 for Total ex-

Table 2 Difference in scores between clinical and non-clinical samples

Measure	Non-clinical		Clinical		Mean Difference	Z (U Mann-Whitney)	p	r	d Cohen
	M	SD	M	SD					
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

* Confidence Intervals of 95%.

cept Risk scores) and within the girls (Z=-4.33 and Z=-4.08 for the Total score and for Total except Risk score, p<0.0005).

The t test showed similar statistically significant differences: ($t_{(92)}$=-3.68, 95% CI from -6.78 to -0.58, p<0.05, for the Total score; $t_{(92)}$=-3.58, 95% CI from -6.82 to -0.35, p<0.05, for the Total except Risk score) both within the boys and ($t_{(139)}$=-5.91, IC 95% CI from -8.26 to -3.55, p<0.05, for the Total score; $t_{(139)}$=-5.46, 95% CI from -7.82 to -3.10, p<0.05, for the Total except Risk score) within the girls. Like in the comparison of the whole samples, and as can be seen in Table 2, discrepancies between the effect sizes appeared, but were even higher for the female sample.

Principal component analysis (PCA)

Parallel analysis, which is a more robust test of number of factors in a dataset than other methods[29], suggested two factors be retained in both samples (clinical and non-clinical). The items mapped to the same factor in both samples, varying only in the precise loadings.

As in the original study[7], the item mappings resulting from the oblique rotation (see Table 3) suggest one component of positively worded (3, 5 and 10) and one of negatively worded items (1, 2, 4, 6, 7, 8 and 9). Moreover, it should be noted that the correlation between factors for the clinical sample was almost identical to that of the original study (0.36).

Table 3 PCA for both samples

Non-clinical sample			Clinical sample		
Items	Factors		Items	Factors	
	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

Cut-off points and Reliable and Clinically significant change

Having found statistically significant differences between sexes, the cut-off points were calculated for both groups separately. In addition, they were also provided for the Total except Risk scale (see Table 4). In relation to the British study of 2015, the results replicated the differences between sexes, with higher cut-off points for girls (almost 3 points above). The cut-off points of the Total except Risk score were slightly higher, a difference that may be attributed to the respondents' tendency to score low in the Risk item.

The estimation of the reliable change was also performed by both sexes separately, using the total sample. In the case of boys, the value of the reliable change was 8.17 for the Total score, and 8.83 for the Total except Risk score, while for girls it was 8.86 for the Total score, and 9.46 for Total except Risk score. Therefore, any change greater than those indicated in each case can be considered a reliable change. In the British article values are given only for the Total score, being 8.2 in boys and 7.7 in girls: very similar to our Spanish values for boys, though in girls the Spanish cutting point is higher by more than one point.

Table 4	Cut-off points for the Total score and for the Total except Risk score	
Group	**CP Total**	**CP Total no R**
Chicos	12.72	13.88
Chicas	15.49	16.64
Muestra Total	14.91	16.05

CONCLUSION

This paper is the first psychometric exploration in Spanish of the YP-CORE, an instrument widely used in the psychological counseling services of the United Kingdom[30] and a measure one meta-analysis suggests is the most reliable test to calculate the change produced by therapy[31].

The results show good psychometric properties for the Spanish version. Firstly, acceptability was good with a high percentage of usable responses, something that can be attributed, amongst other things, to brevity of the YP-CORE. In addition, as in the British study of 2015, it has shown good internal consistency in both samples, as well as for both sexes separately. It also showed acceptable temporal stability, though with a statistically significant reduction in scores over time with low effect sizes, something that was observed in the validation of the original CORE-OM[32] and is not uncommon with psychopathology measures. The convergent validity analysis showed high correlations between almost all the scales considered, but mainly between the overall scores of both questionnaires, and between the YP-CORE Total score and the YSR Internalizing problems broadband. This result suggest greater suitability of the YP-CORE to measure internalised than externalised distress. We also found higher scores for females; the difference in the influence of age between this study and the British one[8] underlines the importance of considering demographic characteristics and doing so for any translation, in addition to the sample size. It is likely that these results reflect socioeconomic, cultural and developmental issues[33]. The initial factorial structure has been replicated by the PCA, reproducing almost exactly the mapping of items by factor the original UK analysis. Similarly, the cutoffs have reflected similarities with recent British results, though they are a little higher for these Spanish participants. Taking this into account, effects of sociodemographic characteristics of samples should be considered, as well as the effect of the medication and diagnoses, comorbidities and severity of symptoms when interpreting scores. As for the estimation of the reliable change, the values identified in the Spanish sample coincide with the British in the case of the boys, but not in the case of the girls. Finally, although the Risk item is the one with the lowest factor loading in both samples, it seems that its exclusion from the Total score does not lead to different results in any of the analysis, so the use of the Total except Risk score does not seem to be statistically necessary.

Though these results strongly support the Spanish YP-CORE, limitations of the present study should be noted. These include a moderately low sample size (both clinical and non-clinical), which makes its generalization difficult and prevented analysis of scores in relation to clinical aspects such as diagnosis or medication. The crucial comparison between the clinical and non-clinical samples showed some overlap: non-clinical participants with slightly elevated scores could fall within the clinical spectrum, suggesting an only moderate discriminative power of the test so it should not be used diagnostically. In addition, the lack of post-therapy data means cannot yet comment on the sensitivity to change of the instrument in clinical contexts, though the reported initial estimates of the reliable change index gives a statistical basis on which to categorise change.

In conclusion, the YP-CORE seems a promising instrument to be used in a wide range of psychological services for young people. The results have reaffirmed its potential to detect psychological distress in adolescent Spanish-speaking populations, as well as to give levels of severity. Its brevity and comprehensibility make it ideal for its routine use in order to monitor the progress and changes produced by the therapeutic intervention. However, studies with larger samples and assessing the effects of therapeutic interventions are still needed.

ACKNOWLEDGEMENTS

In particular, the voluntary and generous participation of Alejandro Botella in the process of translating the test into Spanish is appreciated. The collaboration of the *Institució Pedagògica Sant Isidor* (IPSI), the *Hospital de Mataró,* and the *Institut de Trastorns Alimentaris* (ITA) is also appreciated. This study has been partially funded by the *Generalitat de Catalunya* (ref. SGR717).

CONFLICT OF INTERESTS

The authors do not report any conflict of interest.

REFERENCES

1. Collishaw S, Maughan B, Goodman R, Pickles A. Time trends in adolescent mental health. J Child Psychol Psychiatr. 2004; 45(8):1350-62.
2. 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
3. Caballo VE, Simón MÁ. Manual de psicología clínica infantil y del

adolescente: trastornos específicos. Madrid: Ediciones Pirámide; 2002.
4. Bados A, García E, Fusté A. Eficacia y utilidad clínica de la terapia psicológica. Int J Clin Health Psychol. 2002;2(3):477-502.
5. 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.
6. del Barrio V. Evaluación psicológica en la infancia y la adolescencia. Madrid: Universidad Nacional de Educación a Distancia; 2003.
7. 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.
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.
9. Kovacs M. Children's depression inventory (CDI2). North Tonawanda, NY: Multi-Health Systems, Inc.; 2011.
10. Spielberger CD, Edwards CD, Montuori J, Lushene R. State-Trait Anxiety Inventory for Children. Palo Alto, CA: Consulting Psychologist Press; 1973.
11. Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatr. 1997;38(5):581-6.
12. Achenbach T. Manual for the youth self-report and 1991 profile. Burlington, VT: University of Vermont Department of Psychiatry; 1997.
13. Achenbach T, Edelbrock C. Child Behavior Checklist: CBCL. Burlington, VT: University of Vermont Department of Psychiatry; 1980.
14. Achenbach T, Rescorla L. ASEBA school-age forms & profiles. Burlington, VT: ASEBA; 2001.
15. 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.
16. Forns M. Evaluación psicológica infantil. Barcelona: Barcanova; 1993.
17. 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.
18. 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.
19. Trujillo A, Feixas G, Bados A, García-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.
20. Hambleton RK, Merenda PF, Spielberger CD. Adapting educational and psychological tests for cross-cultural assessment. Tilburg: Psychology Press; 2005.
21. Evans C. Translating and "normalising" CORE system: CORE System Trust (CST) position statement; 2008. Disponible en: http://www.psyctc.org/CORE-OM/Translating_CORE-OM_CST_position_statement.pdf
22. Michael B, Harold PA, Allen F. DSM-IV TR. Manual diagnóstico y estadístico de los trastornos mentales. Barcelona: Masson; 2002.
23. Achenbach T. Manual for the youth self-report and 1991 profile. Burlington, VT: University of Vermont Department of Psychiatry; 1991.
24. Gómez-Gómez M, Danglot-Banck C, Vega-Franco L. Sinopsis de pruebas estadísticas no paramétricas. Cuándo usarlas. Rev Mex Pediatr. 2003;70(2):91-9.
25. Guàrdia J, Peró M, Freixa M, Turbany J. Análisis de datos en psicología (2ª ed.). Barcelona: Delta Publicaciones; 2008.
26. IBM Corporation. IBM SPSS statistics for Windows, Version 21.0. Armonk, NY: IBM Corp; 2012.
27. Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. Michigan: Houghton Mifflin; 2003.
28. 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.
29. Horn J. A rationale and test for the number of factors in factor analysis. Psychometrika. 1965;30(2):179-85.
30. 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.
31. Cooper M. Counselling in UK secondary schools: A comprehensive review of audit and evaluation data. Couns Psychother Res. 2009;9(3):137-50.
32. 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.
33. Greig A, Taylor J, MacKay T. Doing research with children: A practical guide. Glasgow: Sage; 2012.