



Internal structure, reliability and cross-cultural validity of the Warwick-Edinburgh Mental Wellbeing Scale in three European populations

Prateek Yadav,^{1,2} Jorge Arias de la Torre ,^{3,4,5} Ioannis Bakolis,¹ Xavier Bartoll,^{6,7} Cristina Casajuana Kogel,⁸ Joan Colom Farran,⁸ Alex Dregan,¹ Carlos Garcia Forero,⁹ Lorena Botella-Juan,⁵ Vicente Martín,^{4,5} Antonio J Molina,^{4,5} Philippe Mortier,^{4,10} Line Nielsen,¹¹ Katherine Perez,^{4,6,7} Beatriz Puertolas,¹⁰ Amy Ronaldson ,¹ Ziggi Santini,¹² Ana Schiaffino,^{13,14} Antoni Serrano-Blanco,^{4,15} Sarah Stewart-Brown,¹⁶ Jose M Valderas,^{17,18,19} Jordi Alonso,^{4,10,20} Gemma Vilagut^{4,10}

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjment-2024-301433>).

For numbered affiliations see end of article.

Correspondence to

Dr Jorge Arias de la Torre, King's College London, London, UK; jorge.arias_de_la_torre@kcl.ac.uk

PY and JAdT are joint first authors.

JA and GV are joint senior authors.

Received 24 October 2024
Accepted 9 February 2025

ABSTRACT

Background The Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS), a questionnaire designed for the assessment of mental well-being, is widely used in different countries and cultures worldwide. However, there is a lack of studies examining its metric performance and measurement invariance across countries.

Objective This study aims to examine the internal structure, reliability and cross-country validity of the WEMWBS in three European populations.

Methods WEMWBS data collected in 2016 from three representative population health surveys from an autonomous region in Spain (Catalonia) and two countries (Denmark and the UK) were used (n=13 940). The mean WEMWBS Scores were compared between populations. The internal consistency (ω coefficients), internal structure (confirmatory factor analyses (CFA) and bifactor exploratory structural equation models), reliability (item response theory models, item and test information functions), and cross-cultural comparability (multigroup CFA) of the WEMWBS were assessed.

Findings Differences in mean scores observed between regions merit further study. The WEMWBS showed high internal consistency across countries ($\omega=0.942$). The unidimensionality of the scale was confirmed overall and for each population. Evidence of reliability and of measurement invariance at the configural, scalar and metric levels was found.

Conclusions and implications The results support the use of the WEMWBS in different cultures to inform the understanding of population well-being in public health and its possible use as an outcome measure in clinical studies.

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Mental well-being is arousing great interest as an outcome measure from a public health and clinical perspective, worldwide.
- ⇒ One of the most widely used tools specifically designed for the assessment of mental well-being is the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS).
- ⇒ However, evidence about its internal structure, that is, if it measures just mental well-being or also other constructs, and reliability is still sparse, and the comparability of the WEMWBS Scores between different cultural contexts remains unclear.

WHAT THIS STUDY ADDS

- ⇒ The results from this study, one of the largest to date, specifically focused on the assessment of mental well-being and based on large and representative samples from three different cultural contexts in Europe (the UK, Denmark and Catalonia), provide additional robust evidence of the primary focus of the WEMWBS on general mental well-being and advocate for its application as a unidimensional measure.
- ⇒ Also, the results reinforce the reliability and validity of the WEMWBS as a tool for assessing mental well-being in these contexts and demonstrate its measurement invariance across them.

measuring mental well-being is complex because mental well-being is not simply a lack of mental health, but a multifactorial combination of many factors, and both hedonic and eudaimonic perspectives must be considered.

In the absence of reliable detectable biomarkers or investigations for the assessment of mental well-being, self-reported questionnaires are an important resource for its measurement as they are able to provide quantitative data on experiences, feelings and functioning.³ One of these questionnaires,

INTRODUCTION

Mental well-being, also known as positive mental health, is a key priority of the WHO and the Organization for Economic Cooperation and Development recommendations,^{1,2} and is gaining increasing relevance from a public health and clinical perspective, worldwide. Despite its increasing relevance,



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY. Published by BMJ Group.

To cite: Yadav P, Arias de la Torre J, Bakolis I, et al. *BMJ Ment Health* 2025;28:1–7.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ As far as we know, this study is the largest cross-cultural validation of the WEMWBS, confirming the cross-cultural validity of the WEMWBS across diverse European countries, supporting the suitability of the scale as a relevant unidimensional outcome measure for the assessment of mental well-being at the population level, for its use in multicultural population studies, clinical settings and in other contexts, at least at the European level.
- ⇒ The availability of a measure with these characteristics allows us to determine and compare the levels of mental well-being in different contexts, determine the effectiveness of public health and clinical interventions to improve it and, consequently, to increase it at the population level.

likely the most widely used, is the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS), a self-report questionnaire to measure mental well-being, encompassing both hedonic well-being (feelings of happiness and satisfaction) and eudaimonic well-being (ie, functioning, purpose in life, personal development and meaningful relationships).⁴ It consists of 14 Likert-type items describing components of mental well-being phrased as positive statements, as opposed to components of mental ill health phrased in the negative.⁴ While higher WEMWBS Scores are associated with lower scores for psychopathology, mental well-being goes beyond the mere absence of mental illness, and it correlates with measures of health-related behaviours and physical capability.⁴ This is a reason why the WEMWBS has been used as an outcome measure in clinical studies and in a wide variety of non-clinical settings, including the commercial and third sectors and education systems. Besides, a previous systematic review and meta-analysis assessing interventions to improve well-being using changes in WEMWBS Score as the main outcome measure found that the 14-item WEMWBS version could be considered, due to its flexibility, a relevant outcome measure in a wide range of medical, public health and social interventions.⁵ This points out the WEMWBS value as an outcome measure in health services settings, but also at the population level to inform the development of public health policies.

Even though the WEMWBS was initially developed as a unidimensional scale,⁶ the factor structure of the measure has been subject to substantial debate, and alternative structures have been explored. Some studies have proposed two-factor models, including hedonic and eudaimonic well-being,⁷ or a three-factor structure including an additional separate social factor.⁸ Recent studies exploring the dimensionality of the WEMWBS have added to the complexity, with a bifactor structure with a general mental well-being factor alongside specific subfactors on hedonia, eudaimonia or personal relationships demonstrating superior fit than one-factor to three-factor models.^{8–9} Bifactor models are particularly advantageous, as they simultaneously account for the general construct and specific variances,¹⁰ and offer a comprehensive and methodologically robust approach to evaluating the dimensionality of the WEMWBS. However, further research is needed to determine whether they offer meaningful improvements in interpreting WEMWBS Scores over simpler structures.

The WEMWBS was originally developed and validated in the UK but has been translated and adapted for use in many other populations across the world. Although the WEMWBS has been extensively validated for its use in individual countries, including

France, the UK, Denmark and Spain,^{4 11 12} to date there is a lack of evidence about its cross-cultural comparability. One of the largest direct cross-country comparisons to date in Europe involved data collected in Denmark, Iceland, Catalonia and the UK.¹³ However, it addressed WEMWBS validity in Denmark only and did not assess cross-cultural validity between the countries.¹³ Besides, a comparison of WEMWBS between Catalonia and Scotland has been performed, which showed significantly different mean WEMWBS Scores between these countries.¹⁴ Differences in social and cultural contexts may influence how individuals interpret and respond to items, potentially altering the scale's dimensionality. It is essential to ensure that measures of pathology and mental well-being are both reliable and cross-culturally valid—that is, equivalent and comparable across diverse cultural and population groups. A lack of comparability can result in misinterpretations of results, thereby diminishing their utility as measures for informing evidence-based decision-making in mental health. Given the limited research on the cross-cultural validity of the WEMWBS, conducting studies using data from several populations is instrumental in determining its relevance as a mental well-being measure.

In this context, this study aims to investigate the internal structure, reliability and cross-cultural validity of the WEMWBS in a secondary data analysis of population health survey data from three European populations.

METHODS

Participants and data

A cross-sectional study based on representative data from population health surveys from Catalonia (Spain), Denmark and England was carried out. Data from Catalonia (Spain) were obtained from the 2016 Catalan Health Interview Survey (ESCA).¹⁵ Anonymised data from ESCA were requested and accessed after signing a data transfer and confidentiality agreement with the Department of Health of the Government of Catalonia. Catalonia is a region of Spain with a population of 7.7 million in 2022. Respondents were non-institutionalised persons in the general population aged 15 years and over selected from the Population Registry of Catalonia of the Catalan Institute of Statistics. The sample was selected through a stratified three-stage random sampling strategy. The computer-assisted interviews were conducted face to face by trained interviewers at the respondents' residences. Data from the two waves of the survey carried out in 2016 were used, with a total sample size of 4818, but only data on individuals 16 years and over with complete WEMWBS data were extracted and used ($n=3651$).¹⁵

The data from Denmark comes from the Danish Mental Health and Wellbeing Survey 2016 ($n=3508$).¹⁶ This survey used a random representative sample of Danish individuals from the Danish Civil Registration System 16 years and older.¹⁷ Denmark had a population of 4 718 756 at the time of the survey. The survey and sampling design were carried out by Statistics Denmark. Individuals were randomly selected and invited by a letter via digital mail, with information about the survey and instructions for the online questionnaire. Two reminder letters (one postal and one digital) were sent to non-responders. A total of 10 250 individuals were sampled, of which 3508 participated in the survey (34.2%). From these participants, 174 were excluded due to missing data in all or some of the WEMWBS items (4.96%).

Data from the UK were taken from the 2016 National Health Survey for England carried out by the Joint Health Surveys Unit of NatCen Social Research and the Research Department of

Epidemiology and Public Health at University College London. Data were requested and accessed through the UK data service, and use of safeguarded data is governed by a legally binding end user licence agreement, which forms part of the registration process. The survey involved a stratified multistage, random probability sample of the non-institutionalised population living in private households in England.¹⁸ Addresses were randomly selected from 531 postcodes to provide a representative sample. The survey was carried out using face-to-face interviews and a self-reported module to be completed by the participant, including the WEMWBS Questionnaire. A total of 8011 individuals 16 years or older participated (response rate=55%), of which 7153 (89.3%) completed the self-reported module. 198 of these participants were excluded due to missing data of all or some of the WEMWBS items (2.8%).

All data collected from the population surveys were anonymised. Confidentiality and privacy requirements were met through complete anonymisation and through users' licence agreements with the corresponding organisations whenever required. Only data from those over 16 years with complete WEMWBS responses were included to ensure comparability between countries (n=13 940).

Outcome measure

The WEMWBS is a self-report questionnaire consisting of 14 Likert-type items about the last 2 weeks, with a 5-point response scale assessing eudaimonic (eg, people's functioning, social relationships, sense of purpose) and hedonic (eg, feelings of happiness) components of mental well-being. A higher score corresponds to a higher level of mental well-being, with scores ranging from 14 to 70.⁴ The WEMWBS shows suitable psychometric properties for the assessment of mental well-being for each of the populations included in this study (the UK, Denmark and Spain).^{4 11–13}

Statistical analyses

Descriptive analyses

Descriptive statistics were calculated including the mean, median, IQR and SD. Kruskal-Wallis tests were carried out to assess possible differences between countries due to the non-normal distribution of the WEMWBS Scores.

Internal structure and internal consistency

To assess the internal structure of the questionnaire, we compared several factor structures based on previous research: (A) A one-factor confirmatory factor analyses for categorical items (iCFA) model;⁶ (B) A two-factor iCFA model, including hedonic and eudaimonic dimensions;⁷ (C) A three-factor iCFA model with hedonic, eudaimonic and social relations factors;⁸ (D) A three-factor model with a slightly different distribution of items within factors;¹⁹ (E) A second-order three-factor iCFA based on the previous best-fitting three-factor structure;¹⁹ (F) A bifactor exploratory structural equation model (bifactor ESEM) with target rotation with a general factor and three specific factors.⁸ All models were estimated for each population and for the whole data set. The weighted least squares estimator with robust adjustment for mean and variance was used.²⁰ To assess and compare the goodness of fit of the different models, the following statistics were calculated: the χ^2 , Comparative Fit Index (CFI >0.95 for good fit), Tucker-Lewis Index (TLI >0.95 for good fit), root mean square error of approximation (RMSEA <0.06 for good fit, <0.08 for acceptable fit) and standardised root mean square residual (SRMR <0.08 for good fit).²¹ Based

on the bifactor ESEM model, several psychometric indices were obtained that collectively help determine whether subscale scores provide unique information above and beyond the total score, and to what extent the total score can be interpreted as a reliable and valid measure of the intended construct, despite the presence of multidimensionality.²² The indicators studied were: the explained common variance (ECV), or the proportion of all common variances explained by each factor. Values of ECV ≥ 0.85 suggest that the scale is sufficiently unidimensional. The ω reliability coefficient for the multidimensional composite total score (ie, the proportion of the total score variance that can be attributed to all common variances), the ω hierarchy for each factor (ω_H) (that is, the proportion of total score variance that can be attributed to each common factor), indicates the degree to which the raw score reflects the target dimension, and the average relative parameter bias (ARPB), that is, relative difference between item loadings from the unidimensional solution and those from the general factor in the bifactor, that is, the truer model. Additionally, Cronbach's α and McDonald's ω were calculated to assess the internal consistency across the items within the WEMWBS.

Item response theory

For an item response theory (IRT) approach, the following weighted polytomous IRT models were fitted: the partial credit model (PCM), the generalised partial credit model (GPCM) and the graded response model (GRM). The model that best fitted the data was selected using the Akaike Information Criterion. Additionally, the item information functions to evaluate the amount of information provided by each item at a given ability level were plotted for each item across countries.²³ Also, test information functions to assess information of the whole test score throughout the latent continuum (θ_i) were plotted for each country and the whole data set.²³ The higher the information value, the greater the precision of the measurement. Score reliability throughout the continuum was estimated as $1 - (1/\text{Information}(\theta_i))$.

Measurement invariance

To assess the correspondence of the WEMWBS across countries, multigroup CFA were carried out to assess the configural (ie, the consistency of the latent structure of the scale across all groups), metric (ie, whether the items are related to the latent trait of the scale in an equivalent way in all groups) and scalar invariances (ie, whether the items show the same expected response across all groups).²⁴ Weighted least squares estimator with robust adjustment for mean and variance was used. To assess the presence of measurement invariance, the following goodness-of-fit statistics were calculated for a one-factor model: χ^2 , CFI, TLI, RMSEA and SRMR.²¹ Comparisons were made between the different invariance levels by calculating the absolute differences in the fit statistics across levels as proposed by Chen.²⁵ Invariance models are considered to provide an adequate fit if changes from one level of invariance to another in CFI and TLI remain below 0.010, below 0.015 in RMSEA and below 0.030 in SRMR.²⁵

All analyses were weighted using the weights derived from the complex sampling strategies for each of the surveys. All analyses were carried out using the Stata V.17 MP and Mplus 8 software packages.

Table 1 Descriptive statistics for the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) by country

Country	N	Mean	SD	95% CI	Median	IQR
Catalonia (Spain)	3651	58.6	8.3	58.3 to 58.8	59	55–65
Denmark	3334	52.2	8.7	51.9 to 52.5	53	47–58
The UK	6955	49.7	8.8	49.5 to 50.0	50	44–56
Total	13 940	52.7	9.4	52.5 to 52.8	54	47–59
Kruskal-Wallis value of p	<0.001					

RESULTS

Descriptive statistics

The demographic characteristics for the data from each region and for the total sample can be found in online supplemental appendix A. The mean WEMWBS Score was highest in Catalonia (M=58.6, 95% CI 58.3 to 58.8), followed by Denmark (M=52.2, 95% CI 51.9 to 52.5) and the UK (M=49.7, 95% CI 49.5 to 50.0). The differences between countries were statistically significant ($p<0.001$) (table 1). Additionally, a forest plot showing mean WEMWBS Scores (and 95% CI) for each country and for the total sample can be found in online supplemental appendix B.

Internal structure and internal consistency

The goodness-of-fit statistics for the different models estimated are shown in online supplemental appendix C, for the total sample and each region. As expected, given the large sample size, the χ^2 statistic was significant for all the models. Results for the unidimensional model initially hypothesised provide adequate CFI and the TLI values above 0.90 for the overall and for each of the three countries included, and good SRMR lower than 0.05. However, the RMSEA values around 0.10 or higher for all populations indicate inadequate fit for the one-factor model according to this indicator. The bifactor ESEM model with target rotation provided the best fit to the data for all populations, with consistently all goodness-of-fit indices over the prespecified cut-off points indicating good (CFI=0.990, TLI=0.977, SRM=0.014 values for the total sample) or adequate (RMSEA=0.075, 90% CI 0.073 to 0.077 in the total sample) fit.

The full standardised factor loadings from the bifactor ESEM model, which provided the best fit to the data, can be found in the table in online supplemental appendix D. Item 8 (*I've been feeling good about myself*) had the greatest standardised factor loading on the general factor, indicating the highest correlation between the item and well-being, in the data from Denmark and from the UK, and the total sample. In Catalonia, item 10 (*I've been feeling confident*) had the greatest loading in the general factor. Loadings for specific factors were generally small to moderate, with values below 0.4 for most items.

Several psychometric indicators were obtained from the bifactor model (online supplemental appendix E), showing very consistent results across populations. ECV for the general factor was 0.86 for the total sample (ranging from 0.82 in UK and Denmark and 0.85 in Catalonia) and ω H values for the general factor over 0.92 for all populations. ARPB values in all populations were around 4%, well below the 10%–15% cut-off indicating low bias introduced when assuming unidimensionality. All these indicators consistently support the predominantly unidimensional structure of the scale across the three countries and justify the interpretation of the total raw score.

The internal consistency estimates Cronbach's α and McDonald's ω coefficients (based on the unidimensional model) for

Table 2 Cronbach's α coefficient and ω coefficient of the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) by country

Country	N	Cronbach's α coefficient*	ω coefficient†
Catalonia	3651	0.934	0.953
Denmark	3334	0.921	0.941
The UK	6955	0.931	0.959
Total	13 940	0.923	0.942

* α must be greater than 0.85 for the scale to be considered reliable (calculated using standardised factor loadings from a weighted factor analysis for categorical items).

† ω based on the unidimensional model must be greater than 0.70 for the scale to be considered reliable (calculated using the weighted variance of each item and of the total items).

the WEMWBS in the overall sample and in each population are provided in table 2. Both α and ω values were over their respective thresholds ($\alpha>0.85$, $\omega>0.70$) for each population and total, indicating high internal consistency and reliability of the WEMWBS across populations.

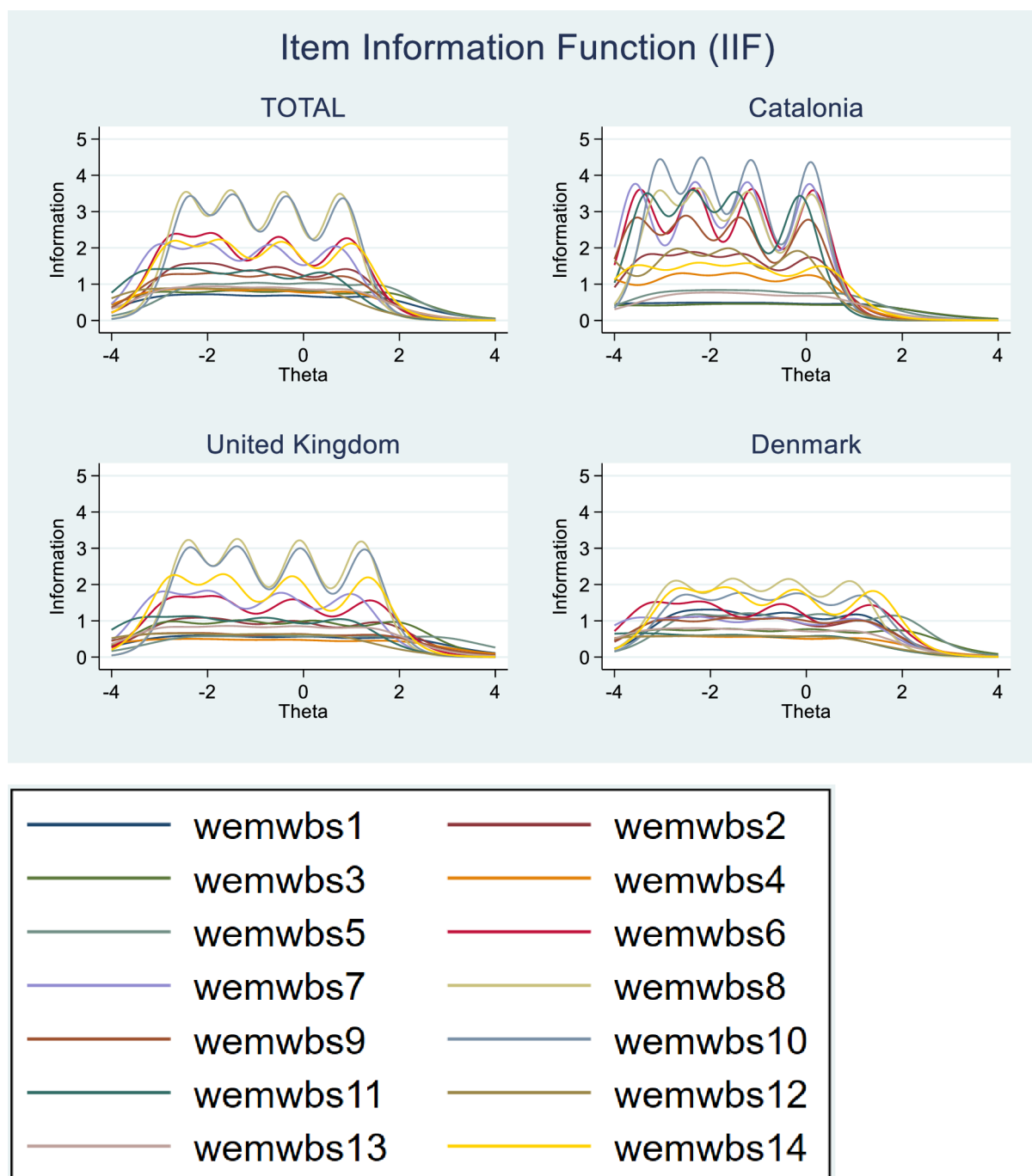
Item response theory

The GRM had the best fit of the different weighted polytomous IRT models, showing the lowest Akaike information criterion, compared with the PCM (total=40.08) and the GPCM (total=39.38). The results of the Akaike information criterion for the different models and for the different populations are provided in online supplemental appendix F. The results from the GRM show that item 8 (*I've been feeling good about myself*) is the item with the greatest ability to discriminate mental well-being in the UK and in Denmark. They also show that in Catalonia item 10 (*I've been feeling confident*) is the one with the greatest discriminating capacity. The item-specific discrimination parameters for the GRM are shown in online supplemental appendix G.

The item information functions are shown in figure 1. In the total sample, items 10 and 8 had the greatest reliability. In Catalonia, item 10 (*I've been feeling confident*) can be seen to have the greatest reliability. In the UK, both items 10 and 8 (*I've been feeling good about myself*) appear to have the greatest reliability. Data from Denmark do not show any item that clearly stands out from the others. However, item 8 appears to have slightly higher reliability compared with other items. Online supplemental appendix H shows the test information function for all countries together and for each country. In all countries, a reliability ≥ 0.90 was observed in the negative part of the continuum of the latent trait. Data from Catalonia had the highest reliability, followed by Denmark, and then the UK.

Measurement invariance

Table 3 shows the fit statistics for configural, metric and scalar invariance for the WEMWBS based on the unidimensional model. All fit statistics for the configural invariance model indicated an adequate fit, except for the RMSEA, which was >0.08 , indicating inadequate fit. This shows that the underlying latent structure is found to be consistent across the three countries, in four out of five of the tests of fit. Also, the metric invariance model, where restrictions were added to test the equality of factor loadings between countries, showed a good fit for all tests with the exception of the RMSEA. Besides, the comparison between the configural and metric invariance models, using differences in the fit statistics CFI, RMSEA and SRMR, also confirmed that the latent structure is consistent across countries ($\Delta\text{CFI}=-0.006$, $\Delta\text{TLI}=-0.013$, $\Delta\text{RMSEA}=0.013$, $\Delta\text{SRMR}=0.000$). Only the difference in the TLI adjustment statistic showed a difference slightly larger than the threshold (-0.013 ,



Note: Labels: wemwbs1: Item 1, wemwbs2: Item 2, wemwbs3: Item 3, wemwbs4: Item 4, wemwbs5: Item 5, wemwbs6: Item 6, wemwbs7: Item 7, wemwbs8: Item 8, wemwbs9: Item 9, wemwbs10: Item 10, wemwbs11: Item 11, wemwbs12: Item 12, wemwbs13: Item 13, wemwbs14: Item 14.

Figure 1 Total item information function (IIF) for each item and for each item by country. WEMWBS, Warwick-Edinburgh Mental Wellbeing Scale.

when the difference should be smaller than 0.010). Finally, when constraints were added to test for equality of thresholds, that is scalar invariance, all model fit statistics indicated a good fit except for the RMSEA, and when comparing the metric invariance model with the scalar invariance model, the differences in all the fit statistics met the criteria proposed by Chen,²⁵ confirming good fit ($\Delta CFI = -0.004$, $\Delta TLI = 0.007$, $\Delta RMSEA = 0.008$, $\Delta SRMR = -0.004$).

DISCUSSION

This study is, as far as we know, the first study to assess over three culturally different populations. The internal structure, reliability

and cross-cultural validity of a tool specifically designed for the assessment of mental well-being, the WEMWBS were assessed. Despite the similarities between the populations (all from Europe), and although there were differences in mean well-being scores across the populations assessed, the WEMWBS showed adequate internal consistency, reliability and cross-country comparability. These findings validate the 14-item version of the WEMWBS as a robust unidimensional measure of mental well-being across diverse settings and cultures, reinforcing its suitability for international use, showing measurement invariance at the European level. These results provide

Table 3 Configural, metric and scalar invariance for the items of the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS)

Invariance	χ^2	CFI	TLI	RMSEA	SRMR
Configural	13 807 087	0.953	0.944	0.112 (0.111 to 0.114)	0.041
Metric	11 981 545	0.959	0.957	0.099 (0.098 to 0.101)	0.041
Scalar	13 281 981	0.955	0.964	0.091 (0.089 to 0.092)	0.045

Note: fit statistics: χ^2

χ^2 with 231 degrees of freedom and $p < 0.001$ for the configural invariance model

χ^2 with 257 degrees of freedom and $p < 0.001$ for the metric invariance model

χ^2 with 339 degrees of freedom and $p < 0.001$ for the scalar invariance model

CFI, Comparative Fit Index; RMSEA, root mean square error of approximation;

SRMR, standardised root mean square residual; TLI, Tucker-Lewis Index.

valuable support for the measurement of well-being and inform the development of international public policies.

Looking at previous studies focused on the WEMWBS, it should be noted that they have been smaller in scope than the present study. A cross-cultural evaluation of the WEMWBS in Chinese and Pakistani origin populations living in the UK found high levels of consistency and reliability; however, this was only in one country.²⁶ Additionally, in line with the results found, a previous study comparing data from Catalonia and Scotland found significantly different WEMWBS Scores and also demonstrated reliability of the WEMWBS in both countries.¹⁴ Besides, to our knowledge, the largest direct cross-country comparison to date of the WEMWBS involved data from Denmark, Iceland, Catalonia and the UK, but it only assessed validity in Denmark.¹³ Thus, while larger studies involving other populations from different countries worldwide are needed, the results found show the potential of the WEMWBS as a mental well-being measure to be used in multicultural studies and contexts.

The difference in mean WEMWBS Scores between countries is consistent with previous research comparing them.¹³ On the other hand, these findings are inconsistent with results from international assessments of happiness and well-being. The World Happiness Report (WHR) and the European Social Survey (ESS) both rated Denmark higher for happiness and well-being than the other regions in this study.^{27 28} This may be due to differences in constructs used for the assessment of well-being. The ESS uses a wide array of questions over many life domains and the WHR used a Cantril Ladder Scale alongside topics as varied as measures of state effectiveness, freedom and social support. A Cantril Ladder involves participants rating these aspects of their life on a scale of 1 to 10. As a result, the ESS and WHR may give a broader view of well-being alongside many other factors (eg, socioeconomic development, political system, environmental factors) and not specifically mental well-being, leading to the difference in results. The WEMWBS provides a more focused measure of mental well-being. To stand out as a tool that can be used both as a population measure and clinical outcome measure needs concise and narrower focus. Furthermore, the conciseness of the WEMWBS lends itself well for use in large-scale population surveys and studies.

The results on the internal structure of the WEMWBS support a strong general well-being factor, consistent with prior studies examining the scale's dimensionality using bifactor models.^{8 9 19} The high values obtained for ECV and the ω_H coefficient and low ARPB underscore the unidimensional interpretability of the scale for mental well-being. As far as we know, only one other study has calculated some of these indicators, obtaining consistent results.¹⁹ The results also show high reliability of the WEMWBS, with minor differences in which items were most discriminatory in each country, which were consistent over different tests. Potential similarities and differences in language, culture and cultural understanding of well-being

might lead to differences in items performing between the different countries. This is a well-described phenomenon for self-reported questionnaires using Likert-type items, as the possible differences in response styles over different countries, and the language and translation used, could affect the responses given.²⁹ Factors affecting responses appear to include country-level features such as extraversion, collectivism and uncertainty avoidance influencing response styles.²⁹ However, these factors were not considered in the surveys included in the study and, hence, further research is required to assess the role of these factors in relation to the WEMWBS performance in different cultures.

Several limitations of the study need to be discussed. First, it only assessed data over three populations from Europe, and this could be expanded to include more populations from other regions worldwide in which mental well-being would be considered conceptually different, such as India or African nations. Second, the cross-sectional nature of the data used for this study and the lack of follow-up of participants preclude testing the WEMWBS validity from a longitudinal perspective. Furthermore, the sample from Catalonia is only from one region of a country, while the other two samples are countrywide samples (Denmark and the UK). Differences in mean scores between different regions could be explored further and could be due to a variety of factors in the three regions. Further analysis in future work adopting a longitudinal perspective could be done regarding the validity of WEMWBS and the differences in its scores between demographic groups, for example, gender and age group. Another potential avenue is exploring the reasons for the similarity of discriminatory capacity and reliability across many items in the WEMWBS, which may indicate redundancy. Indeed, the original authors of the WEMWBS developed a Short Warwick Edinburgh Mental Wellbeing Scale (SWEMWBS) including only seven items.³⁰ However, it should be noted that they acknowledge that the SWEMWBS may present a narrower and more restricted view of well-being than the original WEMWBS and the SWEMWBS questions skew more towards eudaimonic and psychological well-being with little coverage of hedonic well-being.³⁰ Additionally, it must be highlighted that the scope of this study is the assessment of the cross-cultural validity of the original and complete WEMWBS, and not the development of a shorter version of the questionnaire. Further research to develop a shorter and cross-culturally valid version of the WEMWBS could be helpful to disentangle these aspects.

To our knowledge, this is the first cross-cultural validation study of the WEMWBS, comparing three different European countries, adding evidence supporting its use and comparability. The results found highlight its potential to be used as a measure of mental well-being in Europe and to make comparisons of mental well-being between countries. Furthermore, it supports the potential of the WEMWBS to be used as an outcome measure in multicultural and multicountry clinical and epidemiological studies in Europe. Despite the fact that further work is needed to assess the cross-cultural validity of the WEMWBS in other European countries and outside of Europe, the results show the validity of using the WEMWBS in surveys and studies across multiple cultures, aiding assessment and potential improvement of population well-being at the population level.

Author affiliations

¹Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

²South London and Maudsley NHS Foundation Trust, London, UK

³Care in Long Term Conditions Research Division, King's College London, London, UK

⁴CIBER Epidemiology and Public Health (CIBERESP), Madrid, Spain

⁵Institute of Biomedicine (IBIOMED), University of Leon, Leon, Spain

⁶Agència de Salut Pública de Barcelona (ASPB), Barcelona, Spain

⁷Institut de Recerca Sant Pau (IR SANT PAU), Barcelona, Spain

⁸Agència de Salut Pública de Catalunya (ASPCat), Barcelona, Spain
⁹Universitat Internacional de Catalunya, Barcelona, Spain
¹⁰Hospital del Mar Research Institute Barcelona, Barcelona, Spain
¹¹University of Copenhagen, København, Denmark
¹²Capital Region of Denmark, Copenhagen, Denmark
¹³Institut Català d'Oncologia, l'Hospitalet de Llobregat, Llobregat, Spain
¹⁴Departament de Salut, Generalitat de Catalunya, Barcelona, Spain
¹⁵Institut de Recerca Sant Joan de Déu, Parc Sanitari Sant Joan de Déu, Barcelona, Spain
¹⁶University of Warwick, Coventry, UK
¹⁷Department of Medicine, National University of Singapore, Singapore
¹⁸Department of Family Medicine, National University Health System, Singapore
¹⁹Centre for Research in Health Systems Performance (CRiHSP) National University Health System, Singapore
²⁰Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain

X Amy Ronaldson @DrRonaldson

Acknowledgements This article represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Contributors PY and JAdIT share the first authorship of this study. GV and JA verified and accessed the data used for the study. GV acted as guarantor. PY, JAdIT, IB, XB, CCK, JCF, AD, CGF, LBJ, VM, AJM, PM, LN, KP, BP, AR, ZS, AS, ASB, SSB, JMV, JA and GV were involved in the conceptualisation of this manuscript. JAdIT and GV carried out the data analysis. PY, JAdIT, IB, XB, CCK, JCF, AD, CGF, LBJ, VM, AJM, PM, LN, KP, BP, AR, ZS, AS, ASB, SSB, JMV, JA and GV interpreted the results from the analyses and gave feedback on them. PY, JAdIT, JA, GV drafted the manuscript. PY, JAdIT, IB, XB, CCK, JCF, AD, CGF, LBJ, VM, AJM, PM, LN, KP, BP, AR, ZS, AS, ASB, SSB, JMV, JA and GV reviewed and approved submission of the final version of the manuscript. JA and GV are last authors

Funding The present work is funded by CIBER – Consorcio Centro de Investigación Biomédica en Red, Epidemiology and Public Health, ISCIII, Ministerio de Ciencia e Innovación (ESPI0521 Mental GPS), Instituto de Salud Carlos III and cofunded by the European Union (grant numbers PI23/00073; PI19/00109), ISCIII-FSE+ (CP21/00078) MSERVET PMortier; Subdirecció General d'Addiccions, VIH, ITS i Hepatitis Viriques, Secretaria de Salut Pública, Departament de Salut, Generalitat de Catalunya, the Medical Research Council (MRC) and Guy's charity. This article represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South LondonPI23 and Maudsley NHS Foundation Trust and King's College London. Departament de Recerca i Universitats, Generalitat de Catalunya (AGAUR 2021 SGR 00624).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Ethical approval for this study was acquired from the Ethics Committee of the Hospital del Mar (2012/9896) and the University of Leon (ETICA-ULE-032–2021).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing is not applicable as no data sets were generated and/or analysed for this study. Data are available upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iDs

Jorge Arias de la Torre <http://orcid.org/0000-0001-6908-9611>
 Amy Ronaldson <http://orcid.org/0000-0001-8369-9168>

REFERENCES

- World Health Organization. World mental health report: transforming mental health for all. 2022.
- McDaid D, Hewlett E, Park A. Understanding effective approaches to promoting mental health and preventing mental illness. In: *OECD Health Working Papers*. 2017.
- Möller HJ. Standardised rating scales in psychiatry: methodological basis, their possibilities and limitations and descriptions of important rating scales. *World J Biol Psychiatry* 2009;10:6–26.
- Tennant R, Hiller L, Fishwick R, et al. The Warwick-Edinburgh mental well-being scale (WEMWBS): Development and UK validation. *Health Qual Life Outcomes* 2007;5:1–13.
- Blodgett JM, Birch JM, Musella M, et al. What Works to Improve Wellbeing? A Rapid Systematic Review of 223 Interventions Evaluated with the Warwick-Edinburgh Mental Well-Being Scales. *Int J Environ Res Public Health* 2022.
- Tennant R, Hiller L, Fishwick R, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health Qual Life Outcomes* 2007;5:63.
- Serrani Azcurra D. Traducción, adaptación al español y validación de la escala de bienestar mental de WARWICK-EDINBURGH en una muestra de adultos mayores argentinos. *Acta Colomb Psicol* 2015;18:79–93.
- Sarasjärvi KK, Elovainio M, Appelqvist-Schmidlechner K, et al. Exploring the structure and psychometric properties of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) in a representative adult population sample. *Psychiatry Res* 2023;328.
- Peitz D, Thom J, Walther L, et al. Validation of the Warwick-Edinburgh Mental Well-Being Scale for the Mental Health Surveillance (MHS) of German adults. *Health Qual Life Outcomes* 2024;22:92.
- Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual Life Res* 2007;16 Suppl 1:19–31.
- Castellví P, Forero CG, Codony M, et al. The Spanish version of the Warwick-Edinburgh mental well-being scale (WEMWBS) is valid for use in the general population. *Qual Life Res* 2014;23:857–68.
- López MA, Gabilondo A, Codony M, et al. Adaptation into Spanish of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) and preliminary validation in a student sample. *Qual Life Res* 2013;22:1099–104.
- Koushede V, Lasgaard M, Hinrichsen C, et al. Measuring mental well-being in Denmark: Validation of the original and short version of the Warwick-Edinburgh mental well-being scale (WEMWBS and SWEMWBS) and cross-cultural comparison across four European settings. *Psychiatry Res* 2019;271:502–9.
- Forero CG, Adroher ND, Stewart-Brown S, et al. Differential item and test functioning methodology indicated that item response bias was not a substantial cause of country differences in mental well-being. *J Clin Epidemiol* 2014;67:1364–74.
- IDESCAT. Catalan health survey. 2016. Available: <https://www.idescat.cat/pub/?id=esca&lang=en> [Accessed 9 Mar 2023].
- Nielsen L, Hinrichsen C, Santini ZI, et al. Måling af mental sundhed: en baggrundsrapport for sporgeskemaundersøgelsen danskernes trivsel 2016. Syddansk Universitet. Statens Institut for Folkesundhed; 2017.
- Pedersen CB. The Danish Civil Registration System. *Scand J Public Health* 2011;39:22–5.
- Statistics N. Health survey for england 2016 methods health survey for England 2016: methods. Office for National Statistics; 2017.
- Shannon S, Breslin G, Prentice G, et al. Testing the factor structure of the Warwick-Edinburgh Mental Well-Being Scale in adolescents: A bi-factor modelling methodology. *Psychiatry Res* 2020;293.
- Li CH. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav Res Methods* 2016;48:936–49.
- Bentler PM. Comparative fit indexes in structural models. *Psychol Bull* 1990;107:238–46.
- Reise SP, Bonifay WE, Haviland MG. Scoring and modeling psychological measures in the presence of multidimensionality. *J Pers Assess* 2013;95:129–40.
- Baker FB. *The basics of item response theory*. 2001.
- Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika* 1965;30:179–85.
- Chen FF. Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Struct Equ Modeling* 2007;14:464–504.
- Taggart F, Friede T, Weich S, et al. Cross cultural evaluation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) – a mixed methods study. *Health Qual Life Outcomes* 2013;11:27.
- Charalampi A, Michalopoulou C, Richardson C. Validation of the 2012 European Social Survey Measurement of Wellbeing in Seventeen European Countries. *Applied Research Quality Life* 2020;15:73–105.
- Helliwell JF, Layard R, Sachs JD. World happiness report. 2017.
- Harzing AW. Response Styles in Cross-national Survey Research. *Int'l Jnl of Cross Cultural Management* 2006;6:243–66.
- Stewart-Brown S, Tennant A, Tennant R, et al. Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. *Health Qual Life Outcomes* 2009;7:15.