Cross-cultural validity of the WHO-5 Well-Being Index and Euthymia Scale: A clinimetric analysis

Cross-cultural validity of the WHO-5 Well-Being Index and Euthymia Scale: A clinimetric analysis

Danilo Carrozzino[a], Kaj Sparle Christensen[b], Chiara Patierno[a], Agnieszka Woźniewicz[c], Stine Bjerrum Møller[d], Ida-Marie T.P. Arendt[d], Yuqun Zhang[e,f], Yonggui Yuan[f], Natsu Sasaki[g], Daisuke Nishi[g], Carmen Berrocal Montiel[h], Sara Ceccatelli[i], Giovanni Mansueto[i,j], Fiammetta Cosci[i,k,*]

[a]*Department of Psychology "Renzo Canestrari", Alma Mater Studiorum University of Bologna, Bologna, Italy*

[b]*Research Unit for General Practice and Section for General Medical Practice, Department of Public Health, Aarhus University, Aarhus, Denmark*

[c] *Department of Geriatrics, Faculty of Health Sciences, Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń, Poland*

[d]*Department of Psychology, University of Southern Denmark*

[e]*School of Nursing, Nanjing University of Chinese Medicine, Nanjing, China*

[f]*Department of Psychosomatics and Psychiatry, Zhongda Hospital, Southeast University, Nanjing, China*

[g]*Department of Mental Health, Graduate School of Medicine, University of Tokyo, Tokyo, Japan*

[h]*Department of Surgical, Medical and Molecular Pathology, and Critical Care Medicine, University of Pisa, Pisa, Italy*

[i]*Department of Health Sciences, University of Florence, via di San Salvi 12, 50135, Florence, Italy*

[j]*Department of Psychology, Sigmund Freud University, Ripa di Porta Ticinese, 77, 20143, Milan, Italy*

[k]*Department of Psychiatry & Neuropsychology, Maastricht University, PO Box 616, Maastricht, the Netherlands*

[*]**Corresponding author:**

Fiammetta Cosci

Department of Health Sciences, University of Florence,

via di San Salvi 12, 50135, Florence, Italy

Phone: +39 055 2757066; e-mail: fiammetta.cosci@unifi.it

ABSTRACT

Background: The assessment of psychological well-being and euthymia represents an emerging issue in clinical psychology and psychiatry. Rating scales and indices such as the 5-item version of the World Health Organization Well-Being Index (WHO-5) and the Euthymia Scale (ES) were developed but insufficient attention has been devoted to the evaluation of their cross-cultural validity. This is the first study using Clinimetric Patient-Reported Outcome Measures (CLIPROM) criteria to assess cross-cultural validity and sensitivity of five different versions of the WHO-5 and ES.

Methods: A multicenter cross-sectional study involving a total of 3,762 adult participants from different European (i.e., Italy, Poland, Denmark) and non-European (i.e., China, Japan) countries was conducted. Item Response Theory models (Mokken and Rasch analyses) were applied.

Results: Mokken coefficients of scalability were found to range from 0.42 to 0.84. The majority of the versions of the WHO-5 fitted the Rasch model expectations. Paired t-tests revealed that the Italian and Danish WHO-5 versions were unidimensional. Person Separation Reliability indices showed that the Polish, Danish, and Japanese ES versions could reliably discriminate between subjects with different levels of euthymia.

Limitations: A convenience sampling was used, thus limiting the generalizability of study findings. In addition, no measures of negative mental health were administered.

Conclusions: WHO-5 can be used in international studies for cross-cultural comparisons since it covers transcultural components of subjective well-being. Findings also suggest that the ES can be used as a cross-cultural screening tool since it entailed the clinimetric property of sensitivity.

Keywords: clinimetrics; cross-cultural; euthymia; psychological well-being; sensitivity; validity.

1. Introduction

In clinical psychology and psychiatry there has been a reappraisal of both psychological well-being and euthymia (Bech, 2012a; Bech, 2015; Carrozzino et al., 2021a; Fava and Bech, 2016; Fava and Guidi, 2020; Guidi and Fava, 2020; Topp et al., 2015). The concept of psychological well-being refers to a self-reflective state of mind characterized by a subjective sense of satisfaction in different life domains (Bech et al., 1996; Bech, 2012a; Parloff et al., 1954; Topp et al., 2015). Euthymia is a complex condition (Fava and Bech, 2016; Fava and Guidi, 2020; Guidi and Fava, 2020), first outlined as integration of psychic forces (Jahoda, 1958), and more recently (Fava and Bech, 2016) refined as a transdiagnostic condition characterized by the absence of affective disorders and the presence of an individual balance in the dimensions of psychological well-being, flexibility, subjective consistency, and resistance to stress (Fava and Guidi, 2020; Guidi and Fava, 2020).

Over the years, specific tools and indices have been developed to assess both psychological well-being and euthymia (Fava and Bech, 2016; Fava and Guidi, 2020; Topp et al., 2015). The 5-item version of the World Health Organization Well-Being Index (WHO-5) is one of the most sensitive and valid patient-reported outcome measures (PROMs) of subjective well-being (Topp et al., 2015) and can be also used as a highly sensitive screening tool for depression (Carrozzino et al., 2021b). The ES was shown to be a comprehensive instrument for assessing euthymia (Carrozzino et al., 2019; Carrozzino et al., 2021a).

WHO-5 (Awata et al., 2007; Faruk et al., 2021; Khosravi et al., 2015; Saipanish et al., 2009; Topp et al., 2015; Zhang et al, 2021) and ES (Carrozzino et al., 2019; Carrozzino et al., 2021a; Guidi et al., 2019; Merlo et al., 2021; Sasaki et al., 2021; Zhang et al, 2021) have been translated into several languages and used in research studies all over the world. Two studies (Jami and Kemmelmeier, 2020; Sischka et al. 2020) evaluated the WHO-5 invariance across different European countries, but their findings failed to support the cross-cultural validity of the instrument. The ES was validated in several languages (Carrozzino et al., 2019; Carrozzino et al., 2021a; Zhang et al. 2021; Sasaki et al. 2021) but cross-cultural comparisons were not performed, implying that little information is currently available on the cross-cultural validity (Hui and Triandis, 1985) of the ES.

In addition, in most of the above studies, particularly those regarding the WHO-5 (Jami and Kemmelmeier, 2020; Sischka et al., 2020), the validation process was conducted based on classical psychometrics. It has been widely demonstrated that traditional psychometric criteria are likely to clash with the complexity of clinical reality, because of their quest for homogeneity of components (Charlson et al., 2022; Cosci, 2021; Fava et al., 2004; Fava, 2022). Clinimetrics was, instead, proposed by Alvan R. Feinstein (Feinstein, 1982; Feinstein, 1983; Feinstein, 1987) to introduce an innovative assessment method for the evaluation of a number of measurement properties and clinical

issues that did not find room in the traditional psychometric model. Such an approach is nowadays the referral for clinical measurements (Fava et al., 2012) and Clinimetric Patient-Reported Outcome Measures (CLIPROM) criteria (Carrozzino et al., 2021c) have been recently proposed to guide the development and validation of self-reported tools. According to the CLIPROM criteria (Carrozzino et al., 2021c), Item Response Theory (IRT) models (i.e., Rasch and Mokken analyses) test whether assessment instruments are valid measures of the same underlying construct across different cultures (Bech, 2004; Bech, 2012b).

In the present study, a clinimetric analysis of the WHO-5 and ES was performed. According to the CLIPROM criteria (Carrozzino et al., 2021c), the cross-cultural validity of five different (i.e., Italian, Polish, Danish, Chinese, Japanese) versions of the WHO-5 and of the ES was tested via Mokken and Rasch analyses to evaluate the extent to which they were valid measures of the same dimensions of psychological well-being and euthymia across different cultures. More specifically, data from five countries were collected to assess whether the WHO-5 and ES reflected a conceptualization of psychological well-being and euthymia that has cross-cultural applicability across European and non-European (i.e., Asian) cultures. The clinimetric sensitivity of the different versions of the WHO-5 and ES was also examined to assess their ability to discriminate between respondents displaying different levels of psychological well-being or euthymia.

On this background, the present study aimed to provide a response to the following research questions: to what extent can the WHO-5 and ES be used to perform a cross-cultural evaluation and comparison of levels of psychological well-being and euthymia? What are their specific clinical applications when used in cross-cultural studies?

2. Methods

2.1. Procedure and participants

This is a multicenter cross-sectional study involving adult participants from different European (i.e., Italy, Poland, Denmark) and non-European (i.e., China, Japan) countries. Data were collected between October 2020 and May 2021 via an online survey. The optimal number of participants to conduct Rasch analyses was determined using methodological recommendations, which suggest a sample size ranging from 250 to 500 individuals (Hagell and Westergren, 2016). Respondents were not screened for psychiatric disorders and there were no specific requirements (i.e., inclusion and exclusion criteria) for study participation, which was voluntary and not compensated. Participants were recruited from the general population and all study data were anonymously treated. The present study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of Florence (Italy), as well as by the Ethical Committee

of the Universities of Southern Denmark (Denmark) and of the University of Tokyo (Japan). All respondents gave their digital informed consent for study participation.

2.2. Measures

The WHO-5, a widely used self-report measure evaluating the degree of subjective well-being, was used (Topp et al., 2015). Each item (e.g., "I have felt cheerful and in good spirits", "I woke up feeling fresh and rested") is positively worded and scored on a 6-point Likert scale with responses ranging from 0 (at no time) to 5 (all of the time). The WHO-5 raw score ranges from 0 to 25 and is conventionally multiplied by 4 to give a percentage score from 0, representing the worst imaginable well-being, to 100, corresponding to the best imaginable state of subjective well-being (Topp et al., 2015). The different versions of the WHO-5 that were used in the present research are available in pdf-format on the WHO-Five website at the following link: https://www.psykiatri-regionh.dk/who-5/who-5-questionnaires/Pages/default.aspx.

The ES is a 10-item self-rating scale originally developed by Fava and Bech (2016). Five items of the ES (e.g., "My daily life is filled with things that interest me") were derived from the WHO-5 (Topp et al., 2015) and allow to assess a state of subjective well-being. The other ES five items (e.g., "I am able to adapt to changing situations", "Most of the time I can handle stress"), which incorporate Jahoda's (1958) conceptualization of balance and integration of psychic forces, were created with the purpose of detecting psychological flexibility, consistency, and resilience (Fava and Bech, 2016). Each item of the original version of the ES is scored dichotomously as 1 (True) or 0 (False), with a total score ranging from 0 to 10 (Fava and Bech, 2016). Following the aim of having a version of the ES with a response format on a Likert scale, which also homogenizes with the one of the WHO-5, a 6-point Likert scale version of the ES with responses ranging from 1 to 6 was also proposed (Carrozzino et al., 2021a). The Likert scale version of the ES was found to entail the clinimetric properties of construct validity and clinimetric sensitivity (Carrozzino et al., 2021a) and was used in the present research. As to the adaptation process of the different versions of the ES, a translation and back-translation procedure was used; further details are provided elsewhere (Carrozzino et al., 2019; Sasaki et al., 2021; Zhang et al., 2021).

2.3. Statistical analyses

Mokken and Rasch analyses were run to evaluate the cross-cultural validity and clinimetric sensitivity of the WHO-5 and ES. The Mokken analysis, which is the non-parametric version of IRT

models (Mokken, 1971), was conducted to assess scalability, testing the extent to which the WHO-5 and ES were statistically sufficient measures of the underlying dimensions of psychological well-being and euthymia. According to Bech (2012b), Loevinger's coefficients (Loevinger, 1947) ranging from 0.30 to 0.39 indicate a just acceptable level of scalability, while a coefficient ≥ 0.40 is a clear indication of the scalability of the rating scale under evaluation. The Mokken analysis was run using Stata statistical software, version 14.2 (Stata Corporation, College Station, TX).

The Rasch analysis, which is the parametric version of IRT models, was conducted using Rasch Unidimensional Measurement Models (RUMM2030) software (Andrich et al., 2010) and to assess the following clinimetric properties:

- Overall fit to the model, which was tested using the chi-square item-trait interaction statistics providing a summary measure of how the rating scales under evaluation conform to the Rasch model expectations (Nielsen et al., 2017; Pallant and Tennant, 2007; Tennant and Conaghan, 2007). A non-significant chi-square probability value indicates a good level of overall fit to the Rasch model (Nielsen et al., 2017; Pallant and Tennant, 2007; Tennant and Conaghan, 2007).
- Construct validity or dimensionality, which was examined to determine whether the different WHO-5 and ES versions were valid indices of the same underlying dimensions of psychological well-being and euthymia across different cultures. Principal Component Analysis of residuals was run to identify the two most different subsets of items (i.e., the most positively and negatively factor-loading items on the first component). Paired *t*-tests were then performed to compare scores on the two subsets of items. If more than 5% of *t*-tests were significant, indices were not considered unidimensional (Nielsen et al., 2017).
- Local dependency, which was investigated evaluating whether the response to one item was dependent on the response to another item after controlling for the underlying construct under examination (Marais and Andrich, 2008). A residual correlation value > 0.20 is the evidence of the presence of local dependency between items (Christensen et al., 2017a; Christensen et al., 2017b; Marais and Andrich, 2008).
- Differential Item Functioning, which was tested to assess whether a certain form of item bias can occur when different groups of subjects (e.g., males and females) respond differently to an individual item despite equal levels of the dimension under evaluation (Christensen et al., 2017b; Hagell and Westergren, 2016).
-Person Separation Reliability Index, which was assessed to estimate the clinimetric sensitivity of the different versions of the WHO-5 and ES (Carrozzino et al., 2021a; Carrozzino et al., 2021c), that is their ability to discriminate among subjects with different levels of the underlying dimensions of psychological well-being and euthymia.

3. Results

3.1. Descriptive statistics

Demographic characteristics (i.e., age, gender, marital status, and education level) of the European and non-European samples under study are reported in Table 1. Descriptive statistics (i.e., mean scores and standard deviation, as well as minimum and maximum total scores of the WHO-5 and ES) are listed in Table 2.

3.2. Overall fit of the different versions of the WHO-5

Model fit statistics of the WHO-5 are reported in Table 3. Model fit statistics of the Italian version of the WHO-5 showed a significant item-trait interaction statistic ($\chi^2 = 102.45$, df = 45, $p < 0.001$), which indicated misfit to the Rasch model. Even after adjusting the sample size, fit to the model was not achieved ($\chi^2 = 81.57$, df = 45, $p < 0.001$). Model fit statistics of the Polish version of the WHO-5 showed a significant item-trait interaction statistic ($\chi^2 = 68.88$, df = 45, $p = 0.01$), which indicated misfit to the Rasch model. The fit was achieved after adjusting the sample size ($\chi^2 = 50.87$, df = 45, $p = 0.25$). Model fit statistics of the Danish version of the WHO-5 showed a significant item-trait interaction statistic ($\chi^2 = 91.49$, df = 45, $p < 0.001$), indicating misfit to the Rasch model. Even after adjusting the sample size, fit to the model was not achieved ($\chi^2 = 68.58$, df = 45, $p = 0.01$). The initial analysis of the Chinese version of the WHO-5 showed a significant item-trait interaction statistic ($\chi^2 = 66.36$, df = 45, $p = 0.02$), which indicated misfit to the Rasch model. The fit was achieved after adjusting the sample size ($\chi^2 = 51.85$, df = 45, $p = 0.22$). Model fit statistics of the Japanese version of the WHO-5 showed a significant item-trait interaction statistic ($\chi^2 = 89.30$, df = 40, $p < 0.001$), which indicated misfit to the Rasch model. The fit was achieved after adjusting the sample size ($\chi^2 = 47.60$, df = 40, $p = 0.19$).

3.3. Overall fit of the different versions of the ES

As to the Italian version of the ES, the initial analysis showed a significant item-trait interaction statistic ($\chi^2 = 236.44$, df = 90, $p < 0.001$), indicating misfit to the Rasch model. Even after adjusting the sample size (Table 4, Analysis 2) and removing misfitting items (Table 4, Analysis 3-4), fit to the model was not achieved. Concerning the Polish version of the ES, the initial analysis showed a significant item-trait interaction statistic ($\chi^2 = 215.00$, df = 90, $p < 0.001$), indicating misfit to the Rasch model. The fit was achieved after the exclusion of misfitting items ($\chi^2 = 84.91$, df = 72, $p = 0.14$). Standardized fit residuals for items (SD = 1.84) and subjects (SD = 1.12) were within acceptable limits. As to the Danish version of the ES, the initial analysis showed a significant item-

trait interaction statistic ($\chi^2 = 343.08$, df = 90, $p < 0.001$), indicating misfit to the Rasch model. Even after adjusting the sample size (Table 4, Analysis 2) and removing misfitting items (Table 4, Analysis 3-4), fit to the model was not achieved. The initial analysis of the Chinese version of the ES revealed a significant item-trait interaction statistic ($\chi^2 = 289.37$, df = 90, $p < 0.001$), indicating misfit to the Rasch model (Table 4, Analysis 1). Ordered response categories were achieved after rescoring all items but without significantly improving the overall fit to the Rasch model (Table 4, Analysis 2). Even after adjusting the sample size (Table 2, Analysis 3) and excluding misfitting items (Table 4, Analysis 4-9), fit to the model was not achieved. Regarding the Japanese version of the ES, the initial analysis showed a significant item-trait interaction statistic ($\chi^2 = 282.00$, df = 90, $p < 0.001$), which indicated misfit to the Rasch model. Even after adjusting the sample size, fit to the model was not achieved ($\chi^2 = 145.21$, df = 90, $p < 0.001$).

3.4. Scalability and Dimensionality of the different versions of the WHO-5

As to the Italian version of the WHO-5, the total score obtained an acceptable scalability (Loevinger's coefficient of homogeneity of 0.61). Testing for dimensionality confirmed that less than 5% of *t*-tests were significant, indicating that the Italian version of the WHO-5 was unidimensional. Mokken analysis of the Polish version of the WHO-5 showed that the total score had an acceptable scalability (Loevinger's coefficient of homogeneity of 0.66). Testing for dimensionality revealed significant *t*-tests outside the critical value of 5%, suggesting that the Polish version of the WHO-5 was multidimensional. Regarding the Danish version of the WHO-5, the total score had an acceptable scalability (Loevinger's coefficient of homogeneity of 0.56). Testing for dimensionality confirmed that less than 5% of *t*-tests were significant, indicating that the Danish version of the WHO-5 was unidimensional. Mokken analysis of the Chinese version of the WHO-5 showed that its total score had an acceptable scalability (Loevinger's coefficient of homogeneity of 0.84). Testing for dimensionality revealed significant *t*-tests outside the critical value of 5%, suggesting that the Chinese version of the WHO-5 was multidimensional. As to the Japanese version of the WHO-5, Mokken analysis showed that the total score had an acceptable level of scalability (Loevinger's coefficient of homogeneity of 0.77). Testing for dimensionality revealed significant *t*-tests outside the critical value of 5%, suggesting that the Japanese version of the WHO-5 was multidimensional.

3.5. Scalability and Dimensionality of the different versions of the ES

As to the Italian version of the ES, Mokken analysis showed that its total score had an acceptable scalability (Loevinger's coefficient of homogeneity of 0.42). Testing for dimensionality revealed significant *t*-tests outside the critical value of 5%, indicating that the Italian version of the ES was multidimensional (Table 4, Analysis 1-4). Mokken analysis of the Polish version of the ES

indicated that its total score had an acceptable scalability (Loevinger's coefficient of homogeneity of 0.50). Testing for dimensionality revealed significant *t*-tests outside the critical value of 5%, suggesting that the Polish version of the ES was multidimensional. As to the Danish version of the ES, the total score had an acceptable scalability (Loevinger's coefficient of homogeneity of 0.44). Testing for dimensionality revealed significant *t*-tests outside the critical value of 5%, suggesting that the Danish version of the ES was multidimensional. The total score of the Chinese version of the ES had an acceptable scalability (Loevinger's coefficient of homogeneity of 0.56). Testing for dimensionality revealed significant *t*-tests outside the critical value of 5%, suggesting that the Chinese version of the ES was multidimensional (Table 4, Analysis 1-6). After the exclusion of items 1-4, 6, and 10, less than 5% of *t*-tests were significant, indicating that the 6-, 5-, and 4-item versions of the ES were unidimensional (Table 4, Analysis 7-9). Mokken analysis of the Japanese version of the ES showed that its total score had an acceptable scalability (Loevinger's coefficient of homogeneity of 0.60). Testing for dimensionality revealed significant *t*-tests outside the critical value of 5%, suggesting that the Japanese version of the ES was multidimensional.

3.6. Local Dependency

Patterns of local dependency between items of the different versions of the WHO-5 are reported in Table 3. No indication of local dependency between items was found.

Patterns of local dependency between items of the different versions of the ES are reported in Table 4. Indication of local dependency between 29 item-pairs was detected with residual correlations > 0.20.

3.7. Differential Item Functioning (DIF)

Indication of DIF is reported in the supplementary material. As to the WHO-5, item 1 ("If I become sad, anxious or angry it is for a short time") of the Chinese version showed a non-uniform DIF for education. There were no indications of uniform or non-uniform DIF for the other versions of the WHO-5. As to the ES, item 10 ("I wake up feeling fresh and rested") of the Italian version showed a uniform DIF for education, while a uniform DIF for age was detected on item 5 ("Most of the time I can handle stress") of the Polish version.

3.8. Person Separation Reliability Index (PSI)

PSI indices of the WHO-5 are reported in Table 3. The Italian, Polish, Chinese, and Japanese versions were found to display a PSI ranging from 0.87 to 0.93, indicating that these measures could reliably discriminate between subjects with different levels of the construct under examination. As to

the Danish version of the WHO-5, the PSI was 0.82, suggesting that the scale could reliably distinguish between different groups but not between different subjects.

PSI indices of the Polish, Danish, and Japanese versions of the ES were found to range from 0.85 to 0.93, indicating that the scale could reliably discriminate between subjects with different levels of the dimension under evaluation (Table 4). PSI indices of the Italian and Chinese versions of the ES ranged from 0.78 to 0.88, suggesting that these measures could reliably distinguish between different groups but not between different subjects.

4. Discussion

This is the first study applying CLIPROM criteria (Carrozzino et al., 2021c) to the process of assessment of cross-cultural validity of self-reported measures of psychological well-being and euthymia. Our findings, particularly those regarding Mokken analysis, indicate that the WHO-5 and ES are valid cross-cultural measures of psychological well-being and euthymia. Mokken coefficients of scalability were found to range from 0.42 to 0.84, suggesting that the several versions of the WHO-5 and ES cover transcultural dimensions. However, caution should be paid when interpreting the findings as chi-square item-trait interaction statistics revealed that most versions of the ES did not fit the Rasch model expectations mainly because of multidimensionality, and a large proportion of misfitting and locally dependent items. According to the International Quality of Life Assessment (IQOLA) research program, one of the largest projects aimed at translating and validating a measure of health-related quality of life (i.e., the SF-36) for its use in cross-national studies (Bullinger et al., 1998; Raczek et al., 1998), Rasch analysis should be regarded as the most appropriate statistical method for testing cross-cultural validity of rating scales. As the authors of the IQOLA project (Raczek et al., 1998; p. 1204) outlined, Rasch models "enable the test developer to examine the equivalence of item calibrations across different samples and contexts, including various cultural-linguistic settings and translations". This assumption is in line with CLIPROM criteria (Carrozzino et al., 2021c) which acknowledge the utility of Rasch analysis for performing a clinimetric evaluation of how universally valid and useful a scale is (Bech et al., 1978).

As to the WHO-5, although there were some fluctuations in the overall fit to the model, no misfitting items were detected and no indication of local dependency between items was found. The Polish, Chinese, and Japanese versions were found to fit the Rasch model expectations and paired *t*-tests also showed that the Italian and Danish versions of the WHO-5 were found to be unidimensional measures evaluating the same underlying traits of psychological well-being. These findings indicate that the WHO-5 covers transcultural components of subjective well-being and that it can be therefore used in international studies to perform a cross-cultural evaluation and comparisons of levels of

psychological well-being (Jami and Kemmelmeier, 2020). Bech (2012a) anticipated these findings, noting that the WHO-5 has cross-cultural applicability since it covers the basic aspects of a subjective state of well-being (i.e., feeling cheerful and in good spirits, calm and relaxed, active and vigorous, fresh and rested when waking up, and interested in day-to-day activities).

Rasch analysis of the ES showed that the 8-item Polish version was the only one fitting the model. The Italian, Danish, Chinese, and Japanese versions had poor overall fit to the Rasch model and included a number of misfitting items. Question 4 ("I try to be consistent in my attitudes and behaviors") was particularly problematic not being related to the underlying construct of euthymia, which is in line with previous studies (Carrozzino et al., 2021a; Sasaki et al., 2021). There is, therefore, a need for a revision of the way in which item 4 is formulated and scored, particularly when the ES is expected to be used for a cross-cultural evaluation of euthymia. Caution should also be paid since uniform DIF cases were detected, one for education on item 10 ("I wake up feeling fresh and rested") of the Italian ES, and the other for age on item 5 ("Most of the time I can handle stress") of the Polish ES. Paired *t*-tests showed that the ES, except the 6-, 5-, and 4-item forms of the Chinese version, was a multidimensional measure. This might show that the underlying construct under evaluation lacks cross-cultural generalizability (van de Vijver and Poortinga, 1982) but it may also be the case that "translations are rarely sufficiently perfected to provide equivalent meaning across languages" (Bracken and Barona, 1991). Following CLIPROM criteria (Carrozzino et al., 2021c) and methodological recommendations for translating and validating health status questionnaires (Bullinger et al., 1998; Gandek and Ware, 1998; Wagner et al., 1998), future research should be designed not only to clarify the extent to which the ES is a valid measure of the same underlying construct of euthymia across different cultures but also to determine whether translation procedures were conducted in a manner that preserved the conceptual meaning of the original items of the ES to make them culturally relevant within each country.

WHO-5 and ES entailed the clinimetric property of sensitivity. Specifically, the Italian, Polish, Chinese, and Japanese versions of the WHO-5 as well as the Danish, Polish, and Japanese versions of the ES displayed an excellent level of clinimetric sensitivity, suggesting their use as cross-cultural screening measures for discriminating between subjects with different levels of psychological well-being and euthymia. This is in line with a number of previous studies showing that the WHO-5 (Garland et al., 2018; Krieger et al., 2014; Low et al., 2021; Topp et al., 2015) and ES (Sasaki et al., 2021; Zhang et al., 2021) were highly sensitive screening tools for depression.

4.1. Limitations

The present findings should be interpreted in light of the fact that a convenience sampling was used, thus limiting the generalizability of results. To further examine the cross-cultural validity of the Italian, Polish, Danish, Chinese, and Japanese versions of the WHO-5 and ES, future studies should be conducted using representative samples of the general populations of the five different countries under evaluation. In addition, no measures of negative mental health were administered. Future studies, making use of rating scales assessing for instance depression or anxiety, are needed to determine the cross-cultural discriminant validity of the WHO-5 and ES. Future studies, using other measures evaluating positive dimensions of mental health such as the Positive Mental Health Scale (Lukat et al., 2016), are also highly encouraged to further assess the cross-cultural validity of the WHO-5 and ES.

5. Conclusion

In today's world where people are widely inter-connected (Beckstead et al., 2008; Boer et al., 2018), easy-to-use, sensitive, and valid measures applicable across languages and usable by researchers and clinicians for a cross-cultural evaluation of psychological well-being and euthymia are warranted (Vaillant, 2012). Clinimetrics (Fava, 2022) and CLIPROM criteria (Carrozzino et al., 2021c) provide a perfect methodological framework for cross-cultural assessment. Based on the present clinimetric analysis, the WHO-5 entailed the clinimetric properties of sensitivity and construct validity and it was found to reflect a culturally sensitive and inclusive definition of psychological well-being which allows for cross-cultural comparisons. This implies that the WHO-5 can be used as a highly sensitive screening tool, as an outcome measure to detect improvements in psychological well-being after treatment, and also as a dimensional measure for assessing and comparing levels of subjective well-being across different cultures. As to the ES, our findings indicate that the Polish, Danish, and Japanese versions displayed an excellent clinimetric sensitivity, thus suggesting to use this clinimetric index as a screening tool. Both WHO-5 and ES can also have a potential application in cross-cultural studies assessing patients undergoing psychological interventions, which are aimed at building euthymia (Guidi & Fava, 2020) and empowering well-being (e.g., Well-Being Therapy) (Fava, 2016). However, given the limited cross-cultural validity of the Likert scale version of the ES, its use can be supplemented by other indices such as the Clinical Interview for Euthymia (Guidi and Fava, 2022) based on the clinimetric principle of incremental validity (Fava et al., 2012; Sechrest, 1963). Future studies are also encouraged to test the cross-cultural sensitivity and validity of the original version of the ES (Fava and Bech, 2016).

Declarations of interest

None

Contributors

Danilo Carrozzino wrote the first draft of the manuscript; Kaj Sparle Christensen conducted the statistical analyses and revised the final version of the paper; Chiara Patierno collected data and revised the final version of the paper; Agnieszka Woźniewicz collected data and revised the final version of the paper; Stine Bjerrum Møller collected data and revised the final version of the paper; Ida-Marie T.P. Arendt collected data and revised the final version of the paper; Yuqun Zhang collected data and revised the final version of the paper; Yonggui Yuan collected data and revised the final version of the paper; Natsu Sasaki collected data and revised the final version of the paper; Daisuke Nishi collected data and revised the final version of the paper; Carmen Berrocal Montiel collected data and revised the final version of the paper; Sara Ceccatelli revised the final version of the paper; Giovanni Mansueto revised the final version of the paper; Fiammetta Cosci designed the research, supervised data collection, collected data, and revised all versions of the paper. All authors have approved the final version of the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

None

References

Allgaier, A.K., Kramer, D., Saravo, B., Mergl, R., Fejtkova, S., Hegerl, U., 2013. Beside the Geriatric Depression Scale: the WHO - Five Well - being Index as a valid screening tool for depression in nursing homes. Int. J. Geriatr. Psychiatry. 28 (11), 1197-1204. https://doi.org/10.1002/gps.3944

Andrich, D., Lyne, A., Sheridan, B., Luo, G., 2010. RUMM 2030. RUMM Laboratory, Perth, Australia.

Awata, S., Bech, P., Yoshida, S., Hirai, M., Suzuki, S., Yamashita, M., Ohara, A., Hinokio, Y., Matsuoka, H., Oka, Y., 2007. Reliability and validity of the Japanese version of the World Health Organization - Five Well - Being Index in the context of detecting depression in diabetic patients. Psychiatry. Clin. Neurosci. 61 (1), 112-119. https://doi.org/10.1111/j.1440-1819.2007.01619.x

Bech, P., Allerup, P., Rosenberg, R., 1978. The Marke - Nyman Temperament Scale. Evaluation of transferability using the Rasch item analysis. Acta. Psychiatr. Scand. 57 (1), 49-58. https://doi.org/10.1111/j.1600-0447.1978.tb06873.x

Bech, P., Gudex, C., Johansen, K.S., 1996. The WHO (Ten) well-being index: validation in diabetes. Psychother. Psychosom. 65 (4), 183-190. https://doi.org/10.1159/000289073

Bech, P., 2004. Modern psychometrics in clinimetrics. Psychother. Psychosom. 73 (3), 134-138. https://doi.org/10.1159/000076448

Bech, P., 2012a. Subjective positive well-being. World. Psychiatry. 11 (2), 105–106. https://doi.org/10.1016/j.wpsyc.2012.05.021

Bech, P., 2012b. Clinical Psychometrics. Wiley, Blackwell, Oxford.

Bech, P., 2015. Clinical assessments of positive mental health, in: Jeste, D.V., Palmer, B.W. (Eds.), Positive Psychiatry: A Clinical Handbook. American Psychiatric Publishing, Arlington, pp. 127-143.

Beckstead, J.W., Yang, C.Y., Lengacher, C.A., 2008. Assessing cross-cultural validity of scales: A methodological review and illustrative example. Int. J. Nurs. Stud. 45 (1), 110-119. https://doi.org/10.1016/j.ijnurstu.2006.09.002

Boer, D., Hanke, K., He, J., 2018. On detecting systematic measurement error in cross-cultural research: A review and critical reflection on equivalence and invariance tests. J. Cross-Cult. Psychol. 49 (5), 713-734. https://doi.org/10.1177/0022022117749042

Bracken, B.A., Barona, A., 1991. State of the art procedures for translating, validating and using psychoeducational tests in cross-cultural assessment. Sch. Psychol. Int. 12 (1-2), 119-132. https://doi.org/10.1177/0143034391121010

Bullinger, M., Alonso, J., Apolone, G., Leplège, A., Sullivan, M., Wood-Dauphinee, S., Gandek, B., Wagner, A., Aaronson, N., Bech, P., Fukuhara, S., Kaasa, S., Ware, J.E., 1998. Translating health status questionnaires and evaluating their quality: the IQOLA project approach. J. Clin. Epidemiol. 51 (11), 913-923. https://doi.org/10.1016/S0895-4356(98)00082-1

Carrozzino, D., Svicher, A., Patierno, C., Berrocal, C., Cosci, F., 2019. The euthymia scale: a clinimetric analysis. Psychother. Psychosom. 88 (2), 119-122. https://doi.org/10.1159/000496230

Carrozzino, D., Christensen, K.S., Mansueto, G., Brailovskaia, J., Margraf, J., Cosci, F., 2021a. A clinimetric analysis of the euthymia, resilience, and positive mental health scales. J. Affect. Disord. 294, 71-76. https://doi.org/10.1016/j.jad.2021.07.001

Carrozzino, D., Christensen, K.S., Cosci, F., 2021b. Construct and criterion validity of patient-reported outcomes (PROs) for depression: a clinimetric comparison. J. Affect. Disord. 283, 30-35. https://doi.org/10.1016/j.jad.2021.01.043

Carrozzino, D., Patierno, C., Guidi, J., Berrocal Montiel, C., Cao, J., Charlson, M.E., Christensen, K.S., Concato, J., De las Cuevas, C., de Leon, J., Eöry, A., Fleck, M.P., Furukawa, T.A., Horwitz, R.I., Nierenberg, A.A., Rafanelli, C., Wang, H., Sonino, N., Fava, G.A., 2021c. Clinimetric criteria for patient-reported outcome measures. Psychother. Psychosom. 90 (4), 222-232. https://doi.org/10.1159/000516599

Charlson, M.E., Carrozzino, D., Guidi, J., Patierno, C., 2022. Charlson Comorbidity Index: A Critical Review of Clinimetric Properties. Psychother. Psychosom. 91 (1), 8-35. https://doi.org/10.1159/000521288

Christensen, K.B., Makransky, G., Horton, M., 2017a. Critical values for Yen's Q 3: Identification of local dependence in the Rasch model using residual correlations. Appl. Psychol. Meas. 41 (3), 178-194. https://doi.org/10.1177/0146621616677520

Christensen, K.S., Oernboel, E., Zatzick, D., Russo, J., 2017b. Screening for depression: Rasch analysis of the structural validity of the PHQ-9 in acutely injured trauma survivors. J. Psychosom. Res. 97, 18-22. https://doi.org/10.1016/j.jpsychores.2017.03.117

Christensen, K.S., Oernboel, E., Nielsen, M.G., Bech, P., 2019. Diagnosing depression in primary care: A Rasch analysis of the major depression inventory. Scand. J. Prim. Health. Care. 37 (1), 105-112. https://doi.org/10.1080/02813432.2019.1568703

Cosci, F., 2021. Clinimetric perspectives in clinical psychology and psychiatry. Psychother. Psychosom. 90 (4), 217-221. https://doi.org/10.1159/000517028

Fava, G.A., 2016. Well-Being Therapy. Treatment Manual and Clinical Applications. Karger, Basel.

Faruk, M., Alam, F., Chowdhury, K., Soron, T., 2021. Validation of the Bangla WHO-5 Well-being Index. Glob. Ment. Health. 8, e26. https://doi.org/10.1017/gmh.2021.26

Fava, G.A., Ruini, C., Rafanelli, C., 2004. Psychometric theory is an obstacle to the progress of clinical research. Psychother. Psychosom. 73 (3), 145-148. https://doi.org/10.1159/000076451

Fava, G.A., Tomba, E., Sonino, N., 2012. Clinimetrics: the science of clinical measurements. Int. J. Clin. Pract. 66 (1), 11-15. https://doi.org/10.1111/j.1742-1241.2011.02825.x

Fava, G.A., Bech, P., 2016. The concept of euthymia. Psychother. Psychosom. 85 (1), 1-5. https://doi.org/10.1159/000441244

Fava, G.A., Guidi, J., 2020. The pursuit of euthymia. World. Psychiatry. 19 (1), 40-50. https://doi.org/10.1002/wps.20698

Fava, G.A., 2022. Forty years of clinimetrics. Psychother. Psychosom. 91 (1), 1-7. https://doi.org/10.1159/000520251

Feinstein, A.R., 1982. T. Duckett Jones memorial lecture. The Jones criteria and the challenges of clinimetrics. Circulation. 66 (1), 1-5. https://doi.org/10.1161/01.CIR.66.1.1

Feinstein, A.R., 1983. An additional basic science for clinical medicine: IV. The development of clinimetrics. Ann. Intern. Med. 99 (6), 843-848. https://doi.org/10.7326/0003-4819-99-6-843

Feinstein, A.R., 1987. Clinimetrics. Yale University Press, New Haven.

Gandek, B., Ware, J.E., 1998. Methods for validating and norming translations of health status questionnaires: The IQOLA Project approach. J. Clin. Epidemiol. 51 (11), 953-959. https://doi.org/10.1016/S0895-4356(98)00086-9

Garland, A.F., Deyessa, N., Desta, M., Alem, A., Zerihun, T., Hall, K.G., Goren, N., Fish, I., 2018. Use of the WHO's Perceived Well-Being Index (WHO-5) as an efficient and potentially valid screen for depression in a low income country. Fam. Syst. Health. 36 (2), 148–158. https://doi.org/10.1037/fsh0000344

Guidi, J., Piolanti, A., Gostoli, S., Schamong, I., Brakemeier, E.L., 2019. Mental pain and euthymia as transdiagnostic clinimetric indices in primary care. Psychother. Psychosom. 88 (4), 252-254. https://doi.org/10.1159/000501622

Guidi, J., Fava, G.A., 2020. The emerging role of euthymia in psychotherapy research and practice. Clin. Psychol. Rev. 82, 101941. https://doi.org/10.1016/j.cpr.2020.101941

Guidi, J., Fava, G.A., 2022. The clinical science of euthymia: A conceptual map. Psychother. Psychosom. 91 (3), 156-167. https://doi.org/10.1159/000524279

Hagell, P., Westergren, A., 2016. Sample Size and Statistical Conclusions from Tests of Fit to the Rasch Model According to the Rasch Unidimensional Measurement Model (RUMM) Program in Health Outcome Measurement. J. Appl. Meas. 17 (4), 416-431.

Hui, C.H., Triandis, H.C., 1985. Measurement in cross-cultural psychology: A review and comparison of strategies. J. Cross-Cult. Psychol. 16 (2), 131-152. https://doi.org/10.1177/0022002185016002001

Jahoda, M., 1958. Current Concepts of Positive Mental Health. Basic Books, New York. https://doi.org/10.1037/11258-000

Jami, W.A., Kemmelmeier, M., 2020. Assessing Well-Being Across Space and Time: Measurement Equivalence of the WHO-5 in 36 European Countries and Over 8 Years. J. Well-Being Assess. 4 (3), 419-445. https://doi.org/10.1007/s41543-021-00042-8

Kellner, R., 1972. Part 2. Improvement criteria in drug trials with neurotic patients. Psychol. Med. 2 (1), 73-80. https://doi.org/10.1017/s0033291700045645

Khosravi, A., Mousavi, S.A., Chaman, R., Kish, M.S., Ashrafi, E., Khalili, M., Ghofrani, M., Amiri, M., Naieni, K.H., 2015. Reliability and validity of the Persian version of the World Health Organization-five well-being index. Int. J. Health. Stud. 1(1), 17-19.

Krieger, T., Zimmermann, J., Huffziger, S., Ubl, B., Diener, C., Kuehner, C., Holtforth, M.G., 2014. Measuring depression with a well-being index: further evidence for the validity of the WHO Well-Being Index (WHO-5) as a measure of the severity of depression. J. Affect. Disord. 156, 240-244. https://doi.org/10.1016/j.jad.2013.12.015

Loevinger, J., 1947. A systematic approach to the construction and evaluation of tests of ability. Psychol. Monogr. 61 (4), i–49. https://doi.org/10.1037/h0093565

Low, KY., Pheh, K.S., Tan, C.S., 2021. Validation of the WHO-5 as a screening tool for depression among young adults in Malaysia. Curr. Psychol. 2021. https://doi.org/10.1007/s12144-021-02152-1

Lukat, J., Margraf, J., Lutz, R., van der Veld, W.M., Becker, E.S., 2016. Psychometric properties of the positive mental health scale (PMH-scale). BMC. Psychology. 4, 8. https://doi.org/10.1186/s40359-016-0111-x

Marais, I., Andrich, D., 2008. Effects of Varying Magnitude and Patterns of Response Dependence. J. Appl. Meas. 9 (2), 105-124.

Merlo, E.M., Stoian, A.P., Motofei, I.G., Settineri, S., 2021. The Role of Suppression and the Maintenance of Euthymia in Clinical Settings. Front. Psychol. 12, 1534. https://doi.org/10.3389/fpsyg.2021.677811

Mokken, R.J., 1971. A theory and procedure of scale analysis. De Gruyter Mouton, The

Netherlands. https://doi.org/10.1515/9783110813203

Nielsen, M.G., Ørnbøl, E., Vestergaard, M., Bech, P., Christensen, K.S., 2017. The construct validity of the major depression inventory: a Rasch analysis of a self-rating scale in primary care. J. Psychosom. Res. 97, 70-81. https://doi.org/10.1016/j.jpsychores.2017.04.001

Pallant, J.F., Tennant, A., 2007. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br. J. Clin. Psychol. 46 (1), 1-18. https://doi.org/10.1348/014466506X96931

Parloff, M.B., Kelman, H.C., Frank, J.D., 1954. Comfort, effectiveness, and self-awareness as criteria of improvement in psychotherapy. Am. J. Psychiatry. 111 (5), 343-352. https://doi.org/10.1176/ajp.111.5.343

Raczek, A.E., Ware, J.E., Bjorner, J.B., Gandek, B., Haley, S.M., Aaronson, N.K., Apolone, G., Bech, P., Brazier, J.E., Bullinger, M., Sullivan, M., 1998. Comparison of Rasch and summated rating scales constructed from SF-36 physical functioning items in seven countries: results from the IQOLA Project. J. Clin. Epidemiol. 51 (11), 1203-1214. https://doi.org/10.1016/S0895-4356(98)00112-7

Saipanish, R., Lotrakul, M., Sumrithe, S., 2009. Reliability and validity of the Thai version of the WHO‐Five Well‐Being Index in primary care patients. Psychiatry. Clin. Neurosci. 63 (2), 141-146.

Sasaki, N., Carrozzino, D. Nishi, D., 2021. Sensitivity and concurrent validity of the Japanese version of the Euthymia scale: a clinimetric analysis. BMC Psychiatry 21, 482. https://doi.org/10.1186/s12888-021-03494-7

Sechrest, L., 1963. Incremental validity: a recommendation. Educ. Psychol. Meas. 23 (1), 153-158. https://doi.org/10.1177/001316446302300113

Sischka, P.E., Costa, A.P., Steffgen, G., Schmidt, A.F., 2020. The WHO-5 well-being index–validation based on item response theory and the analysis of measurement invariance across 35 countries. J. Affect. Disord. Rep. 1, 100020. https://doi.org/10.1016/j.jadr.2020.100020

Tennant, A., Conaghan, P.G., 2007. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis. Care. Res. 57 (8), 1358-1362. https://doi.org/10.1002/art.23108

Topp, C.W., Østergaard, S.D., Søndergaard, S., Bech, P., 2015. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother. Psychosom. 84 (3), 167-176. https://doi.org/10.1159/000376585

Vaillant, G.E., 2012. Positive mental health: is there a cross-cultural definition? World. Psychiatry, 11 (2), 93-99. https://doi.org/10.1016/j.wpsyc.2012.05.006

van de Vijver, F.J., Poortinga, Y.H., 1982. Cross-cultural generalization and universality. J. Cross-Cult. Psychol. 13 (4), 387-408. https://doi.org/10.1177/0022002182013004001

Wagner, A.K., Gandek, B., Aaronson, N.K., Acquadro, C., Alonso, J., Apolone, G., Bullinger, M., Bjorner, J., Fukuhara, S., Kaasa, S., Leplège, A., Sullivan, M., Wood-Dauphinee, S., Ware, J.E., 1998. Cross-cultural comparisons of the content of SF-36 translations across ten countries: Results from the IQOLA Project. J. Clin. Epidemiol. 51 (11), 925-932. https://doi.org/10.1016/s0895-4356(98)00083-3

Zhang, Y., Wang, X., Carrozzino, D., Yuan, Y., 2021. Clinimetric properties of the Chinese version of the Euthymia scale. Clin. Psychol. Psychother. 29 (1), 360-366. https://doi.org/10.1002/cpp.2592

Table 1

Demographic characteristics of the samples under study

	European countries (*N* = 2,082)			Non-European countries (*N* = 1,680)	
	Italian sample (*N* = 636)	Polish sample (*N* = 728)	Danish sample (*N* = 718)	Chinese sample (*N* = 680)	Japanese sample (*N* = 1,000)
Age M ± SD	34.75 ± 12.26	29.30 ± 10.96	34.85 ± 12.39	39.13 ± 10.63	40.55 ± 11.18
(Min-Max)	(18-74)	(18-75)	(18-76)	(18-76)	(21-65)
Gender *N* (%)					
Male	111 (17.5%)	172 (23.6%)	182 (25.3%)	219 (32.2%)	500 (50%)
Female	523 (82.2%)	554 (76.1%)	532 (74.1%)	459 (67.5%)	497 (49.7%)
Other (e.g., transgender)	2 (0.3%)	2 (0.3%)	3 (0.4%)	2 (0.2%)	3 (0.3%)
Missing	0 (0%)	0 (0%)	1 (0.1%)	0 (0%)	0 (0%)
Marital status *N* (%)					
Single	238 (37.4%)	306 (42%)	184 (25.6%)	121 (17.8%)	462 (46.2%)
Married	213 (33.5%)	199 (27.3%)	433 (60.3%)	498 (73.2%)	435 (43.5%)
Non-marital relationship	150 (23.6%)	207 (28.4%)	101 (14.1%)	25 (3.7%)	10 (1%)
Separated	15 (2.4%)	2 (0.3%)	0 (0.0%)	3 (0.4%)	1 (0.1%)
Divorced	18 (2.8%)	10 (1.4%)	0 (0.0%)	32 (4.7%)	89 (8.9%)
Widower	2 (0.3%)	4 (0.5%)	0 (0.0%)	1 (0.1%)	3 (0.3%)
Education level *N* (%)					
Primary school	0 (0%)	4 (0.5%)	10 (1.4%)	1 (0.1%)	11 (1.1%)
Secondary school	50 (7.9%)	3 (0.4%)	175 (24.3%)	20 (2.9%)	229 (22.9%)
High school	221 (34.7%)	317 (43.5%)	375 (52.1%)	36 (5.3%)	262 (26.2%)
Graduation	365 (57.4%)	404 (55.5%)	123 (17.1%)	510 (75%)	493 (49.3%)

Other	0 (0%)	0 (0%)	37 (5.14%)	113 (16.6%)	5 (0.5%)

M = Mean; SD = Standard Deviation; Min = Minimum; Max = Maximum

Table 2

Descriptive statistics of WHO-5 and ES total scores

	European countries (*N* = 2,082)						Non-European countries (*N* = 1,680)			
	Italian sample (*N* = 636)		Polish sample (*N* = 728)		Danish sample (*N* = 718)		Chinese sample (*N* = 680)		Japanese sample (*N* = 1000)	
	M ± SD	**Min-Max**	**M ± SD**	**Min-Max**	**M ± SD**	**Min-Max**	**M ± SD**	**Min-Max**	**M ± SD**	**Min-Max**
WHO-5	11.64 ± 4.95	0-24	12.78 ± 6.03	0-25	14.54 ± 4.63	2-24	11 ± 6.26	0-25	13.27 ± 5.76	0-25
ES	40.40 ± 8.93	15-60	39.29 ± 9.67	10-60	41.13 ± 8.86	17-60	47.56 ± 9.26	10-60	36.14 ± 9.46	10-60

WHO-5 = 5-item version of the World Health Organization Well-Being Index; ES = Euthymia Scale; M = Mean; SD = Standard Deviation; Min = Minimum; Max = Maximum

Table 3

Model fit statistics for WHO-5 scale items

Sample	Analysis	Model fit (overall)	Item fit residual, mean (SD)	Person fit residual, mean (SD)	PSI with extremes	Dimensionality, significant *t*-tests (CI 95)%	Local dependency (Residual Correlations >0.20)
Italian sample (n=628)	1	$\chi 2(45)=102.45$, $p<0.001$	0.29 (2.58)	-0.41 (1.71)	0.87	3.82 (2.12-5.53)	None
-adjusted sample (n=500)	2	$\chi 2(45)=81.57$, $p<0.001$	0.29 (2.58)	-0.41 (1.71)	0.87	3.82 (2.12-5.53)	-
Polish sample (n=677)	1	$\chi 2(45)=68.88$, $p=0.01$	0.66 (2.87)	0.32 (1.52)	0.87	6.20 (4.56-7.85)	None
-adjusted sample (n=500)	2	$\chi 2(45)=50.87$, $p=0.25$	0.66 (2.87)	0.32 (1.52)	0.87	6.20 (4.56-7.85)	-
Danish sample (n=667)	1	$\chi 2(45)=91.49$, $p<0.001$	-0.64 (2.65)	0.52 (1.34)	0.82	3.60 (1.94-5.25)	None
-adjusted sample (n=500)	2	$\chi 2(45)=68.58$, $p=0.01$	-0.64 (2.65)	0.52 (1.34)	0.82	3.60 (1.94-5.25)	-
Chinese sample (n=640)	1	$\chi 2(45)=66.36$, $p=0.02$	-2.19 (2.95)	-0.65 (3.13)	0.93	7.05 (5.39-8.70)	None
-adjusted sample (n=500)	2	$\chi 2(45)=51.85$, $p=0.22$	-2.19 (2.95)	-0.65 (3.13)	0.93	7.05 (5.39-8.70)	-

Japanese sample (n=938)	1	χ2(40)=89.30, p<0.001	-1.78 (4.19)	0.24 (2.56)	0.92	6.40 (5.00-7.79)	None
-adjusted sample (n=500)	2	χ2(40)=47.60, p=0.19	-1.18 (4.19)	0.24 (2.56)	0.92	6.40 (5.00-7.79)	-

χ2: chi-square; p: probability; SD: standard deviation; PSI: person separation index (with extremes)

Table 4

Model fit statistics for Euthymia scale items

Sample	Analysis	Model fit (overall)	Item fit residual, mean (SD)	Person fit residual, mean (SD)	PSI	Dimensionality, significant *t*-tests (CI95) %	Local dependency (Residual Correlations >0.20)
Italian sample (n=630)	1	$\chi 2(90)=236.44$, $p<0.001$	0.32 (3.06)	0.50 (0.91)	0.86	11.53 (10.42-13.91)	Item 6&7, 9&10
-adjusted sample (n=500)	2	$\chi 2(90)=187.65$, $p<0.001$	0.32 (3.06)	0.50 (0.91)	0.86	11.53 (10.42-13.91)	Item 6&7, 9&10
-delete item 6	3	$\chi 2(81)=121.17$, $p=0.003$	0.33 (2.11)	0.47 (0.86)	0.83	9.00 (9.94-13.31)	Item 9&10
-delete item 4	4	$\chi 2(82)=100.45$, $p=0.02$	0.59 (2.25)	0.36 (0.90)	0.83	8.06 (8.29-12.11)	Item 9&10
Polish sample (n=718)	1	$\chi 2(90)=215.00$, $p<0.001$	0.51 (3.34)	0.45 (1.13)	0.89	8.40 (6.90-10.09)	Item 6&7
-adjusted sample (n=500)	2	$\chi 2(90)=149.71$, $p<0.001$	0.51 (3.34)	0.45 (1.13)	0.89	8.40 (6.90-10.09)	Item 6&7
-delete item 6	3	$\chi 2(81)=107.02$, $p=0.03$	0.54 (2.27)	0.40 (1.06)	0.87	7.30 (5.79-8.98)	None
-delete item 4	4	$\chi 2(72)=84.91$, $p=0.14$	0.60 (1.84)	0.30 (1.12)	0.87	7.99 (6.48-9.67)	None
Danish sample (n=668)	1	$\chi 2(90)=343.08$, $p<0.001$	0.34 (4.08)	0.63 (0.98)	0.87	11.01 (9.43-12.73)	Item 1&2, 8&10
-adjusted sample (n=500)	2	$\chi 2(90)=256,80$ $p<0.001$	0.34 (4.08)	0.63 (0.98)	0.87	11.01 (9.43-12.73)	Item 1&2, 8&10
-delete item 4	3	$\chi 2(81)=158.39$ $p<0.001$	0.57 (3.72)	0.62 (1.10)	0.87	11.46 (9.81-13.11)	Item 1&2

-delete item 6	4	$\chi 2(72)=107.21$ p=0.005	0.63 (2.77)	0.53 (1.03)	0.85	9.23 (7.58-10.87)	None
Chinese sample (n=626)	1	$\chi 2(90)=289.37$, p<0.001	0.22 (5.08)	1.50 (1.65)	0.88	13.15 (11.40-14.79)	Item 6&7, 6&8
-recode (012234)	2	$\chi 2(90)=276.92$, p<0.001	-0.05 (4.87)	1.64 (1.76)	0.88	12.11 (10.47-13.75)	Item 6&7, 6&8
-adjusted sample (n=500)	3	$\chi 2(90)=221.18$, p<0.001	-0.05 (4.87)	1.64 (1.76)	0.88	12.11 (10.47-13.75)	Item 6&7, 6&8
-delete item 2	4	$\chi 2(81)=187.53$, p<0.001	-0.37 (4.84)	1.81 (1.89)	0.88	8.71 (7.07-10.36)	Item 6&7
-delete item 1	5	$\chi 2(64)=162.61$, p<0.001	-1.01 (3.77)	2.00 (2.05)	0.88	8.57 (6.93-10.21)	Item 6&7
-delete item 4	6	$\chi 2(56)=135.01$, p<0.001	-1.22 (4.12)	2.04 (2.16)	0.88	6.50 (4.86-8.14)	Item 6&7
-delete item 3	7	$\chi 2(48)=144.31$, p<0.001	-1.48 (3.60)	2.20 (2.34)	0.88	3.99 (2.35-5.63)	Item 6&7
-delete item 10	8	$\chi 2(40)=144.31$, p<0.001	-1.62 (3.81)	2.36 (2.33)	0.84	2.81 (1.16-4.45)	None
-delete item 6	9	$\chi 2(28)=53.08$, p=0.003	-0.86 (1.87)	2.12 (2.13)	0.78	4.58 (2.94-6.22)	None
Japanese sample (n=971)	1	$\chi 2(90)=282.00$ p<0.001	0.01 (4.60)	0.04 (1.67)	0.93	10.90 (9.55-12.25)	Item 1&2, 6&7, 6&8

-adjusted sample (n=500)	2	$\chi2(90)=145.21$ p<0.001	0.01 (4.60)	0.04 (1.67)	0.93	10.90 (9.55-12.25)	Item 1&2, 6&7, 6&8

$\chi2$: chi-square; p: probability; SD: standard deviation; PSI: person separation index (with extremes)