ELSEVIER

CrossMark

Journal of Clinical Epidemiology 67 (2014) 1364−1374

Journal of Clinical Epidemiology

Differential item and test functioning methodology indicated that item response bias was not a substantial cause of country differences in mental well-being

Carlos G. Forero[a,c], Núria D. Adroher[a,c], Sarah Stewart-Brown[b], Pere Castellví[a,c], Miquel Codony[d], Gemma Vilagut[a,c], Anna Mompart[d], Ricard Tresseres[d], Joan Colom[d], José I. Castro[e], Jordi Alonso[a,c,f,*]

[a]*CIBER Epidemiología y Salud Pública (CIBERESP), Instituto de Salud Carlos III, Mechor Fernandez de Almagro, 3-5, 28029, Madrid, Spain*
[b]*Department of Statistics and Epidemiology, Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK*
[c]*Health Services Research Group, IMIM—Institut Hospital del Mar d'Investigacions Mèdiques, Carrer del Doctor Aiguader, 88, PRBB building, 08003 Barcelona, Spain*
[d]*Agència de Salut Pública de Catalunya, Subdirecció General de Drogodependències, Carrer Roc Boronat 81-95, 08005 Barcelona, Spain*
[e]*Institute of Neuropsychiatry and Addictions (INAD), Parc de Salut Mar, Passeig Maritim 25-29, 08003 Barcelona, Spain*
[f]*Universitat Pompeu Fabra (UPF), Plaça de la Mercè, 10-12. 08002 Barcelona, Spain*

Accepted 23 June 2014; Published online 21 August 2014

Abstract

Objectives: Establishing the cross-cultural equivalence of the mental well-being construct, as measured with the Warwick-Edinburg Mental Well-being Scale (WEMWBS), by studying potential construct validity biases in two countries with previously reported score differences.

Study Design and Setting: We compared the WEMWBS total scores and item responses in Scotland (N = 779) and Catalonia (N = 1,900) general population samples. To assess whether the questionnaire spuriously favored higher scores in Catalonia, we tested for differential item functioning (DIF) by applying ordinal logistic regression on Item Response Theory scores. DIF was tested with likelihood ratio tests and standard effect measures (McFadden Pseudo R^2, >0.13; relative parameter change, >5%), and differential test functioning (DTF) was tested by plotting differences between full-test and purified (i.e., without DIF items) score estimates.

Results: Catalonia showed higher levels of mental well-being than Scotland (Cohen d = 0.84). Three of 14 WEMWBS items showed small amounts of DIF. DIF did not accrue to DTF, as shown by intraclass correlation coefficient (ICC, 0.999) and case-by-case differences (maximum, 0.12 SD) between total and purified scores. Population differences remained mainly constant across sociodemographics and health outcomes.

Conclusion: The WEMWBS measures a distinct well-being construct that is stable across countries, implying that Scotland and Catalonia populations are effectively different in the distribution of mental well-being. This result adds to previous psychometric information and supports WEMWBS as a valid unbiased measures for individual and cross-cultural comparisons.

Keywords: Bias; Cross-cultural studies; Measurement invariance; Construct validity; Patient-reported outcomes measurement; Health outcomes; Positive mental health; Inequalities; Positive psychology; Happiness

1. Introduction

Mental well-being is gaining importance for evaluating health policies [1]. Researchers and governments recognize the need to promote good mental health, beyond the presence of psychopathology [2,3]. However, mental well-being is a construct that might vary across countries, beliefs, and ethnicities due to the cultural construction of welfare [4−6]. Thus, public health researchers and mental health practitioners are in need of

Conflict of interest: All authors declare that they have neither conflicts of interest nor relevant financial disclosure to inform.

Funding: This work was supported by grants from Instituto de Salud Carlos III (CDI12/00440). Additional support included a grant from DIUE of Generalitat de Catalunya (2009 SGR 1095 and 2014 SGR 748) and the support of Agència de Salut Pública de Catalunya, Subdirecció General de Drogodependències, Generalitat de Catalunya, Spain.

* Corresponding author. Tel.: 00 34 93 316 07 60; fax: 00 34 93 316 07 97.
E-mail address: jalonso@imim.es (J. Alonso).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2014.06.017

What is new?

- The large differences in average mental well-being scores based on the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) between Scotland and Catalan populations were not attributable to different cultural structures in the mental well-being construct. Moreover, these differences remained almost constant when segmenting by sociodemographic group.

What it adds?

- The mental well-being construct has been considered liable to differ across cultures. Our results show that is not necessarily so, even when there are large differences between populations under study. The WEMWBS has shown to be a valid instrument for cross-national comparisons.

What should change?

- Cross-country studies of well-being based on self reports should routinely apply DIF techniques in order to discard unobserved, nuisance variables acting as sources of differences between populations. This approach permits to explore the presence of such variables, which would avoid confounding effects when studying the determinants of varying well being distributions.

robust well-being measures that allow fair cross-cultural comparisons [7].

There are a number of well-known instruments that focus on mental well-being, such as the Positive and Negative Affect Scale [8] or Satisfaction with Life Scale [9]. A more recent questionnaire, the Warwick-Edinburg Mental Well-being Scale (WEMWBS), developed and validated in English in a number of population samples [10,11] has attracted international attention. The WEMWBS measures mental well-being, understood as positive experiencing, feeling good, and functioning well. The WEMWBS is a unidimensional measure of mental well-being during the previous 2 weeks, composed of 14 positively phrased items [12] each with a five-point ordered categorical Likert scale, and it is scored from low to high mental well-being as a direct item sum score (minimum, 14 and maximum 70). It shows good psychometric properties, low standard error of measurement, and good responsiveness to change [13], but its cultural stability has been less studied. The English version of the WEMWBS has shown minor psychometric differences when addressed to different minority ethnic groups, namely UK citizens of Chinese or Pakistani ascent [14]. However, these were people living in the United Kingdom who responded to the questionnaire in English and among whom the younger age groups could be expected to have reasonable levels of acculturation in the United Kingdom.

Recently, the WEMWBS was adapted and validated into Spanish [10,15] in student and population samples in Catalonia. The adaptation process was scrupulous and led to a version that showed similar results regarding factor structure and fit, scale reliability, and correlations with external measures [10]. Strikingly, the score distributions of the population of Scotland, United Kingdom, and Catalonia, Spain, were quite different in shape, with Catalonia having a much higher average score (mean, 58.1) than Scotland (mean, 50.0). Up to date, it is not known whether this result reflects real differences in mental well-being across both countries or, rather, it is an indication of language or cultural differences in the understanding of mental well-being. Although the first possibility would imply real national well-being inequalities, the second leaves open the possibility that score comparisons will be unfair because of some form of bias.

1.1. Detecting bias and real differences across populations using differential item and test functioning methodologies

In the case of cross-cultural studies, these differences between score groups might be related with factors not addressed by the questionnaire. Other influences, such as language or cultural issues (familial structure, religious beliefs and attitudes, and socioeconomic country differences) and the like, can be affecting individual responses. An inquiry about the presence of such influences in country scores can be addressed from differential item functioning (DIF) methodology.

The DIF procedures aim at determining the discrepancies between item responses in groups—in our case, country—indicated by a grouping variable [16] $G = \{1,2\}$ (in our case, Scotland and Catalonia). Using any measurement model, item responses depend on target trait Θ (e.g., well-being) of interest, but they may also receive causal influences from external unspecified dimensions. A theoretical framework for the causes of DIF is given by the multidimensional theory of DIF [17]: These dimensions are not measured nor identified (e.g., cultural dimensions) but are differently distributed in each group grouping variable. If DIF exists in an item u_j, respondents with matched trait values Θ choose category k with different probability depending on group membership, that is

$$P_j\left(u_j = k \middle| \Theta, G = 1\right) \neq P_j\left(u_j = k \middle| \Theta, G = 2\right). \quad (1)$$

In our case, it would imply that individuals with the same well-being level have different probability of endorsing a response option in each country. Conversely, DIF would not exist if the groups under study have the same probability of response, so that