Psychological Medicine

cambridge.org/psm

Original Article

Cite this article: Jaya ES *et al* (2022). Comparing psychotic experiences in low-and-middle-income-countries and high-income-countries with a focus on measurement invariance. *Psychological Medicine* **52**, 1509–1516. https://doi.org/10.1017/S0033291720003323

Received: 29 December 2019
Revised: 10 July 2020
Accepted: 26 August 2020
First published online: 7 October 2020

Key words:
Cross-cultural; cross-culture; cross-national; psychosis; schizophrenia; validity

Author for correspondence:
Edo S. Jaya,
E-mail: edo.jaya@ui.ac.id

Comparing psychotic experiences in low-and-middle-income-countries and high-income-countries with a focus on measurement invariance

Edo S. Jaya[1], Caroline Wüsten[2], Behrooz Z. Alizadeh[3,4], Therese van Amelsvoort[3,5], Agna A. Bartels-Velthuis[3,4], Nico J. van Beveren[3,6,7,8], Richard Bruggeman[3,4,9], Wiepke Cahn[3,10,11], Lieuwe de Haan[3,12,13], Philippe Delespaul[3,5], Jurjen J. Luykx[3,10,14], Inez Myin-Germeys[3,15], Rene S. Kahn[3,10,16], Frederike Schirmbeck[3,12,13], Claudia J. P. Simons[3,5,17], Neeltje E. van Haren[3,10,18], Jim van Os[3,10,19], Ruud van Winkel[3,5,15], Eduardo Fonseca-Pedrero[20,21], Emmanuelle Peters[22,23], Hélène Verdoux[24], Todd S. Woodward[25,26], Tim B. Ziermans[27] and Tania M. Lincoln[2]

[1]Psychosis Studies Research Group, Faculty of Psychology, Universitas Indonesia, Depok, Indonesia; [2]Clinical Psychology and Psychotherapy, Institute of Psychology, Universität Hamburg, Hamburg, Germany; [3]Genetic Risk and Outcome of Psychosis (GROUP) Investigators; [4]University of Groningen, University Medical Center Groningen, University Center for Psychiatry, Rob Giel Research Center, The Netherlands; [5]Department of Psychiatry and Neuropsychology, Maastricht University Medical Center, School for Mental Health and Neuroscience, Maastricht, The Netherlands; [6]Antes Center for Mental Health Care, Rotterdam, The Netherlands; [7]Department of Psychiatry, Erasmus MC, Rotterdam, The Netherlands; [8]Department of Neuroscience, Erasmus MC, Rotterdam, The Netherlands; [9]Department of Clinical and Developmental Neuropsychology, University of Groningen, Groningen, The Netherlands; [10]Department of Psychiatry, University Medical Center Utrecht, Brain Centre Rudolf Magnus, Utrecht University, Utrecht, The Netherlands; [11]Altrecht, General Mental Health Care, Utrecht, The Netherlands; [12]Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands; [13]Arkin, Institute for Mental Health, Amsterdam, The Netherlands; [14]Department of Translational Neuroscience, University Medical Center Utrecht, Brain Center Rudolf Magnus, Utrecht, The Netherlands; [15]Department of Neuroscience, Research Group Psychiatry, KU Leuven, Leuven, Belgium; [16]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [17]GGzE Institute for Mental Health Care, Eindhoven, The Netherlands; [18]Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC, Rotterdam, The Netherlands; [19]Department of Psychosis Studies, King's College London, King's Health Partners, Institute of Psychiatry, London, UK; [20]Department of Educational Sciences, University of La Rioja, La Rioja, Spain; [21]Centro de Investigación Biomédica en Red de Salud Mental, Oviedo, Spain; [22]Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [23]South London and Maudsley NHS Foundation Trust, London, UK; [24]University Bordeaux, U1219 Bordeaux Population Health Research Center, Bordeaux, France; [25]Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada; [26]BC Mental Health and Addictions Research Institute, Vancouver, BC, Canada and [27]Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

Abstract

Background. The prevalence of psychotic experiences (PEs) is higher in low-and-middle-income-countries (LAMIC) than in high-income countries (HIC). Here, we examine whether this effect is explicable by measurement bias.

Methods. A community sample from 13 countries ($N = 7141$) was used to examine the measurement invariance (MI) of a frequently used self-report measure of PEs, the Community Assessment of Psychic Experiences (CAPE), in LAMIC ($n = 2472$) and HIC ($n = 4669$). The CAPE measures positive (e.g. hallucinations), negative (e.g. avolition) and depressive symptoms. MI analyses were conducted with multiple-group confirmatory factor analyses.

Results. MI analyses showed similarities in the structure and understanding of the CAPE factors between LAMIC and HIC. Partial scalar invariance was found, allowing for latent score comparisons. Residual invariance was not found, indicating that sum score comparisons are biased. A comparison of latent scores before and after MI adjustment showed both overestimation (e.g. avolition, $d = 0.03$ into $d = -0.42$) and underestimation (e.g. magical thinking, $d = -0.03$ into $d = 0.33$) of PE in LAMIC relative to HIC. After adjusting the CAPE for MI, participants from LAMIC reported significantly higher levels on most CAPE factors but a significantly lower level of avolition.

Conclusion. Previous studies using sum scores to compare differences across countries are likely to be biased. The direction of the bias involves both over- and underestimation of PEs in LAMIC compared to HIC. Nevertheless, the study confirms the basic finding that PEs are more frequent in LAMIC than in HIC.

Introduction

The prevalence of psychotic disorders (psychosis) has traditionally been thought to be similar across countries and cultures (Jablensky, 2000). However, a review of studies indicates clearly that prevalence rates vary across countries (McGrath, Saha, Chant, & Welham, 2008). The evidence from this review also suggests that both incidence and prevalence of psychotic disorders are lower in low-and-middle-income-countries (LAMIC) than in high-income countries (HIC) (McGrath et al., 2008), albeit more comparative research is necessary to further substantiate these findings. Paradoxically, the prevalence of psychotic experiences (PEs) at a subclinical level of psychosis, that are often considered to be an indicator of psychosis proneness (van Os, 2016), has been shown to be higher in LAMIC than in HIC (McGrath et al., 2015) and it is unclear to date why this is the case.

Some researchers have argued that the paradox occurs due to cultural differences. People in LAMIC have been assumed to have better social support [e.g. collective culture in LAMIC (Jablensky et al., 1992)], to be less distressed by PEs (Wüsten et al., 2018), and to perceive PEs as culturally acceptable (Al-Issa, 1995; Luhrmann, Padmavati, Tharoor, & Osei, 2015) rather than as an indicator of poor mental health (McGrath et al., 2016). Together, these factors are assumed to increase the likelihood of reporting PEs and to reduce symptom-related distress. Although this is plausible, it is necessary to rule out another explanation first, namely that the higher rates of PEs in LAMIC are due to measurement invariance (MI, DeVylder & Koyanagi, 2018). If this were the case, people in LAMIC are interpreting the questions used in typical self-report assessments of PEs differently and we would expect the prevalence difference in PEs in LAMIC *v.* HIC to be significantly smaller after correcting for MI. In order to rule out the MI explanation, it is necessary to compare the frequency of PEs across countries using an instrument that has shown to be MI (Putnick & Bornstein, 2016). MI can be tested statistically at four different levels: *Configural MI* refers to equal factor structure of the measurement instrument across groups. *Metric MI* refers to equal factor loadings across groups. *Scalar* MI refers to equal intercepts of indicators across groups. *Residual* MI refers to the equal residual variance of the indicators across groups. Scalar MI is the absolute minimum requirement for comparing scores between groups (van de Schoot, Lugtig, & Hox, 2012). Complete MI is given if MI is shown at all four levels.

Previous studies investigating MI in psychotic symptom measures have produced mixed results, both for PEs in general and for specific symptoms, such as paranoia and hallucinations. For example, Fonseca-Pedrero et al. (2018) utilized the brief version of Schizotypal Personality Questionnaire (SPQ-B) to examine MI across samples from 14 countries from different continents. Although they found evidence for a universal existence of the schizotypy construct across countries, they could not find evidence for metric MI indicating that the contribution of the individual items to the schizotypy construct varied across countries. Similarly, Cicero, Martin, and Krieg (2019*b*) reported non-invariance for items related to positive symptoms of the Wisconsin Schizotypy Scales in a US sample of Asian, White, Hispanic, and Multiracial undergraduate students. In contrast, a dichotomized version of the Community Assessment of Psychic Experiences (CAPE) was shown to have MI in a sample of people with psychotic disorder, siblings, and healthy controls from Brazil, France, Italy, the Netherlands, Spain and the UK (Pignon et al., 2018). Furthermore, a study on the dichotomized version of Prodromal Questionnaire–Brief in a US sample of Asian, White, Hispanic, and Multiracial undergraduate students in two public universities (Cicero, Krieg, & Martin, 2019*a*) was MI whereas the continuum distress scale was not. In terms of specific symptom domains, a study investigating the MI of the Psychosis Screening Questionnaire (PSQ) across samples from different ethnic groups living in the UK showed that paranoid symptoms were measured less validly among Pakistani ethnic groups in comparison to White British ethnic groups and that this exaggerated the ethnic differences in the reported frequency of paranoid symptoms (Heuvelman, Nazroo, & Rai, 2018). In the same study, hallucination symptoms were found to be MI, which is in line with another cross-national study using the Launay-Slade Hallucination Scale-Extended (Siddi et al., 2019). In contrast, in a sample of psychology students from the Netherlands, Nigeria, and Norway the hallucination factor of the positive symptom factor of the CAPE was found to be non-invariant whereas the paranoia and bizarre experiences factors were found to have partial scalar MI (Vermeiden et al., 2019). Despite some promising findings, the overall picture of these studies points to problems when it comes to validly interpreting score differences in psychosis measures between different countries or ethnic groups.

In the present study, we tested the MI of the CAPE (Stefanis et al., 2002), a self-report tool that is frequently used to assess PEs in cross-country comparisons, across country income groups (LAMIC *v.* HIC). The sample was a large multinational community sample drawn from individuals living in 13 countries—Canada, Colombia, France, Germany, Ghana, India, Indonesia, Mexico, Spain, Sweden, the Netherlands/Belgium, the UK, and the USA. Finally, we compared the CAPE scores across country income groups taking putative MI into account to test the hypothesis that differences of PEs across country income groups are not due to MI.

Method

Participants and procedure

The analyses were based on a large data pool of community samples from 13 countries ($N = 7141$), which has been described in detail elsewhere (Wüsten et al., 2018). The data were compiled from 2016 to 2017. It consisted of data from previously published studies identified in a review article of studies using the CAPE (Mark & Toulopoulou, 2016) and from additional data sources. We obtained community sample data from Canada, France, Germany, the Netherlands/Belgium, Sweden, Spain, and the UK. Details of the data collection procedure including ethical approvals are presented in the respective studies (Fonseca-Pedrero, Paino, Lemos-Giráldez, & Muñiz, 2012; Korver et al., 2012; Peters et al., 2016; Schlier, Jaya, Moritz, & Lincoln, 2015; Verdoux, Sorbara, Gindre, Swendsen, & van Os, 2003; Ziermans, 2013). Additional data were collected via online surveys for the purpose of this study. To this aim, participants were recruited using a crowdsourcing website (Crowdflower), on which users participate in surveys for financial compensation. Participants from Colombia, Ghana, India, Indonesia, Mexico, and the USA were recruited via this website only whereas participants from Canada, Germany, and the UK were recruited via this website to complement the pre-existing data. All participants in the online surveys provided written informed consent and indicated being above 18 years of age prior to data collection. The

Ethics Committee at the Universität Hamburg approved the survey.

Of the 7141 participants, 553 were from India (7.7%), 607 from Colombia (8.5%), 185 from Ghana (2.6%), 568 from Indonesia (8.0%), 559 from Mexico (7.8%), 681 from France (9.5%), 839 from Sweden (11.7%), 516 from the Netherlands (7.2%), 658 from Spain (9.2%), 1225 from Germany (17.2%), 257 from the UK (3.6%), 216 from the USA (3.0%), and 277 participants were from Canada (3.9%). There were 3676 male participants (51.5%) and the average age was 27.39 years (S.D. = 9.97), ranging from 18 to 80 years. Demographic characteristics and average scores of the factors of the CAPE by country are shown in Table 1.

The country income groups were formed according to World Bank criteria (World Bank, 2016). Participants from Colombia, Ghana, India, Indonesia, and Mexico were combined to form a LAMIC group and participants from Canada, France, Germany, the Netherlands/Belgium, Spain, Sweden, the UK, and the USA were combined to form a HIC group.

Measures

Community assessment of psychic experiences

The CAPE is a self-report measure of lifetime PEs consisting of 42 items that include positive symptoms (20 items, e.g. 'Do you ever hear voices when you are alone?'), negative symptoms (14 items, e.g. 'Do you ever feel that you experience few or no emotions at important events?') and depressive symptoms (8 items, e.g. 'Do you ever feel pessimistic about everything?', Stefanis et al., 2002). Each item is answered in regard to the frequency on a 4-point Likert scale ('never', 'sometimes', 'often', 'nearly always') and in regard to distress ('not distressed', 'a bit distressed', 'quite distressed', 'very distressed'). The original factorial structure of the CAPE consists of three factors (positive symptoms, negative symptoms, and depression; Stefanis et al., 2002) and the most recently published factorial structure of the CAPE has 11 factors (Schlier et al., 2015), that includes the three main factors called positive symptoms, negative symptoms, and depression, five factors that load into positive symptoms factor called bizarre experiences, paranoia, hallucination, magical thinking, and grandiosity factors, and three factors that load into negative symptoms factor called social withdrawal, blunted affect, and avolition factor.

In this study, the frequency scale of the CAPE was administered in paper-and-pencil ($n = 3792$) and online formats ($n = 3349$). The psychometric properties of the French, English, German, Indonesian, Spanish and Swedish version of the CAPE have been independently examined (Brenner et al., 2007; Fonseca-Pedrero et al., 2012; Jaya, 2017; Schlier et al., 2015; Ziermans, 2013).

Analyses

As a preliminary analysis, we computed a series of confirmatory factory analyses (CFA) to confirm that the 11 factors factorial structure of the CAPE (Schlier et al., 2015) had a better fit to the complete sample data than the three factors (Stefanis et al., 2002) and the proposed seven factors factorial structure (Mark & Toulopoulou, 2016).

Then, we used a series of multiple group CFA (MGCFA) to examine the MI of the CAPE across the two country income groups with increasingly stringent criteria for equivalence to establish configural, metric, scalar, and residual MI (van de Schoot et al., 2012). Where the invariance constraints were found to not meet the proposed criteria, modification indices were used to identify parameters that needed to be freed (e.g. by taking into account residual correlations, or factor loadings of items) across groups in order to test for partial MI.

First, we tested for configural MI by computing MGCFA across the two groups according to the best fitting CAPE factorial structure found in the preliminary CFA. Here we tested whether the same latent constructs (e.g. positive symptoms factor, negative symptoms factor and depression factors) and their compositions (e.g. 20 items load on the positive symptoms factor) could be identified in LAMIC and HIC. Second, we tested for metric MI by repeating the first analysis with the additional assumption that factor loadings are equal across groups. Here we tested whether the answers of participants from LAMIC and HIC reflect the same understanding of the latent constructs (e.g. is the paranoia factor an indicator of the positive symptoms factor, it the conspiracy item an indicator of paranoia factor in both LAMIC and HIC samples). Third, we tested for scalar MI by repeating the MGCFAs with the assumption that loadings AND intercepts are equal across the two groups. Thus, here we tested whether the factor loadings AND the levels of the underlying items (i.e. the intercepts) are equal across groups. In this third step, the factor means were fixed to zero in one group and estimated in the other, while the common factor variances were freely estimated. This produced factor means, which were fixed to zero in all groups in the first and second step. Fourth, we tested for residual MI by repeating the MGCFA with the assumption that loadings, intercepts, AND residuals are equal across groups. For example, the conspiracy item is residual MI if it satisfied all the conditions specified above AND had equal error variance across LAMIC and HIC.

The CFAs and MGCFAs were conducted with structural equation modeling (SEM) using the lavaan package ver. 0.5–23 (Rosseel, 2012) in R version 3.4.3. The missing data were dealt with by pairwise deletion in the MI analyses and listwise deletion in the other analyses. The data were not mean-centered. All models were estimated using diagonally weighted least squares using covariance matrices, which is recommended for ordinal data. The following fit indices, along with the proposed cut-off criteria, were used to assess model fit: CFI >0.95, RMSEA <0.06, and SRMR <0.08 (Hu & Bentler, 1999). The χ^2 is reported but not used as a fit criterion because it tends to reject models that are based on large sample size (Bentler & Bonett, 1980). To compare the fit of three factorial structures, we used the chi-square difference test following the Satorra-Bentler formula to compare nested models (Satorra & Bentler, 2001). To assess MI, we followed the recommended stringent criteria from a simulation study because of our large sample size ($N > 300$), i.e. non-invariant is indicated by Δ CFI $\geqslant$0.10, supplemented by Δ RMSEA $\geqslant$0.15 or Δ SRMR $\geqslant$0.30 (Chen, 2007). We report the analysis of MI following a recommended convention (Putnick & Bornstein, 2016). However, because unequal sample sizes between groups have been shown to reduce MI (Yoon & Lai, 2018), we also report an additional MI analysis on an equalized number of participants between the LAMIC and HIC group by randomly excluding the HIC participants to match the LAMIC number of participants. Finally, we computed *Cohen's d* based on the pooled standard deviation of the groups in order to compare the means of the CAPE factors across groups. Cohen's d can be interpreted as small ($d = 0.20$), medium ($d = 0.50$), and large ($d = 0.80$) differences (Cohen, 1988).

https://doi.org/10.1017/S0033291720003323 Published online by Cambridge University Press

https://doi.org/10.1017/S0033291720003323 Published online by Cambridge University Press

Table 1. Participants' characteristics ($N = 7141$)

Characteristic	India	Colombia	Ghana	Indonesia	Mexico	France	Sweden	The Netherlands/ Belgium	Spain	Germany	UK	USA	Canada	Total
Sex, no. of males (%)	407 (73.6%)	489 (80.6%)	175 (94.6%)	447 (78.7%)	414 (74.1%)	69 (10.1%)	389 (46.4%)	257 (49.8%)	194 (29.5%)	512 (41.8%)	111 (43.2%)	86 (39.8%)	126 (45.5%)	3676 (51.5%)
Test administration, no. administrated online (%)	553 (100%)	607 (100%)	185 (100%)	568 (100%)	559 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	291 (23.8%)	172 (66.9%)	216 (100%)	198 (71.5%)	3349 (46.9%)
Age, mean (s.d.)	28.86 (9.30)	26.61 (9.29)	24.61 (7.14)	29.23 (8.41)	26.80 (8.63)	19.97 (3.01)	26.05 (5.04)	35.82 (13.01)	20.31 (2.63)	26.59 (8.90)	40.42 (13.36)	34.35 (12.30)	34.86 (12.14)	27.39 (9.97)
Positive symptoms factor	1.00 (0.48)	0.88 (0.46)	1.03 (0.42)	0.80 (0.43)	0.86 (0.47)	0.49 (0.28)	0.34 (0.25)	0.28 (0.18)	0.25 (0.22)	0.48 (0.39)	0.45 (0.47)	0.74 (0.62)	0.54 (0.44)	0.58 (0.46)
Bizarre experiences factor	1.01 (0.55)	0.77 (0.58)	0.93 (0.5)	0.72 (0.5)	0.73 (0.58)	0.32 (0.32)	0.22 (0.29)	0.12 (0.19)	0.11 (0.22)	0.29 (0.44)	0.33 (0.5)	0.61 (0.7)	0.42 (0.53)	0.46 (0.53)
Hallucination factor	0.67 (0.68)	0.43 (0.57)	0.51 (0.65)	0.47 (0.56)	0.43 (0.61)	0.12 (0.24)	0.05 (0.19)	0.04 (0.13)	0.05 (0.24)	0.15 (0.42)	0.2 (0.51)	0.45 (0.75)	0.23 (0.46)	0.25 (0.50)
Paranoia factor	1.10 (0.55)	1.05 (0.55)	1.09 (0.5)	0.96 (0.51)	1.08 (0.56)	0.83 (0.43)	0.54 (0.35)	0.51 (0.29)	0.54 (0.33)	0.84 (0.45)	0.75 (0.55)	0.95 (0.65)	0.78 (0.51)	0.82 (0.51)
Magical thinking factor	0.96 (0.68)	0.98 (0.75)	1.18 (0.68)	0.94 (0.66)	0.86 (0.78)	0.96 (0.71)	0.42 (0.56)	0.58 (0.65)	0.3 (0.49)	0.51 (0.65)	0.48 (0.64)	0.80 (0.80)	0.60 (0.63)	0.69 (0.71)
Grandiosity factor	1.48 (0.76)	1.64 (0.79)	2.09 (0.78)	1.19 (0.72)	1.62 (0.78)	0.55 (0.61)	0.81 (0.67)	0.42 (0.54)	0.36 (0.57)	0.84 (0.68)	0.65 (0.82)	1.19 (0.85)	0.91 (0.77)	0.98 (0.84)
Negative symptoms factor	1.11 (0.55)	1.13 (0.57)	0.96 (0.5)	1.02 (0.42)	1.18 (0.56)	0.79 (0.37)	0.88 (0.44)	0.85 (0.27)	0.64 (0.34)	0.97 (0.45)	0.90 (0.56)	1.02 (0.64)	0.90 (0.49)	0.94 (0.49)
Social withdrawal factor	1.11 (0.63)	1.15 (0.67)	0.95 (0.55)	1.09 (0.50)	1.22 (0.64)	0.82 (0.52)	1.02 (0.52)	0.99 (0.40)	0.67 (0.45)	1.03 (0.54)	1.00 (0.66)	1.09 (0.73)	0.99 (0.59)	1.00 (0.58)
Blunted affect factor	1.08 (0.64)	1.04 (0.68)	0.98 (0.62)	0.97 (0.53)	1.04 (0.68)	0.58 (0.49)	0.61 (0.60)	0.66 (0.45)	0.42 (0.50)	0.75 (0.67)	0.70 (0.65)	0.79 (0.71)	0.68 (0.59)	0.77 (0.64)
Avolition factor	1.13 (0.61)	1.16 (0.61)	0.97 (0.56)	1.01 (0.47)	1.22 (0.60)	0.86 (0.41)	0.93 (0.51)	0.85 (0.33)	0.71 (0.36)	1.03 (0.48)	0.94 (0.61)	1.07 (0.70)	0.94 (0.56)	0.98 (0.53)
Depressive symptoms factor	1.05 (0.54)	0.95 (0.56)	0.82 (0.46)	0.97 (0.45)	0.96 (0.55)	1.05 (0.46)	0.81 (0.45)	0.82 (0.31)	0.69 (0.35)	0.94 (0.46)	0.98 (0.62)	1.00 (0.66)	0.93 (0.56)	0.92 (0.49)

Note. The sex distribution (χ^2 (12) = 1405.81, $p < 0.01$) and average age (F (12, 7128) = 207.84, $p < 0.01$) differed significantly across countries.

Table 2. Fit indices of the confirmatory factor analyses

Model	Satorra-Bentler χ^2			CFI	RMSEA (90% CI)	SRMR	AIC	Comparison
	χ^2	*df*	*p*					
Total sample (*N* = 7141)								11<7*<3*
Three factors	13 745.80	816	<0.01	0.852	0.056 (0.055–0.056)	0.059	563 250.51	
Seven factors	10 346.61	810	<0.01	0.891	0.048 (0.047–0.049)	0.057	558 599.43	
Eleven factors	9071.42	808	<0.01	0.905	0.045 (0.044–0.045)	0.055	556 873.87	

Note. * $p < 0.05$. CFI = Robust Comparative Fit Index. RMSEA = Robust Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. AIC = Akaike Information Criterion. 90% CI = 90% Confidence Interval. Comparison = Nested model comparison using a Scaled Chi-Square Difference Test with Satorra-Bentler method.

Results

Confirming the factorial structure of the CAPE across countries

The fit indices of the three, seven, and 11 factors factorial structures of the total sample are presented in Table 2. All tested factorial structures met two out of three cut-off criteria for acceptable fit indicating that all had an acceptable fit. However, the 11 factors factorial structure had a significantly better fit than the other tested factorial structures. Similar results were found when we compared factorial structures within each country, except for the sample from the UK and Ghana in which the 11 factors factorial structure did not have a significantly better fit than the other factorial structures (see online Supplementary Table S1). Factor loadings of the 11 factors factorial structure in the total sample are presented in online Supplementary Table S2.

Measurement invariance of the CAPE across country income groups

The fit indices of the results of MI analyses of the CAPE across country groups are presented in Table 3. We found evidence for configural MI; the three fit indices met the cut-off criteria. We found evidence for metric MI as the ΔCFI, RMSEA, and SRMR met the cut-off criteria. We did not find evidence for full scalar MI because ΔCFI was above the cut-off criterium. Then, we inspected the modification indices and found that freeing seven parameters consisting of factor loadings (i.e. the two factors grandiosity and hallucination of the positive symptoms factor, the two items 'influenced by devices' and ' thought echo' of the bizarre experiences factor, and the item 'being special' of the grandiosity factor) and item intercepts (i.e. lack of hygiene" and ' influenced by devices') resulted in partial scalar MI. Residual MI was not found because the ΔCFI was above the cut-off criteria. Similar results were found in the sample, in which the number of participants was equalized across country groups. However, the scalar and residual MI models achieved goodness of fit for MI (online Supplementary Table S3).

Recomputing the MI of the CAPE across the country groups with the adjusted factorial structure (i.e. excluding the non-invariant parameters) we found evidence for configural, metric, and scalar MI. This allowed us to conduct a non-biased latent score comparison, which is reported in Table 4. This comparison indicated that the participants from LAMIC had significantly higher scores on most CAPE factors (e.g. bizarre experiences, $d = 0.54$, $p < 0.05$; affective flattening, $d = 0.75$, $p < 0.05$; positive symptoms, $d = 0.83$, $p < 0.05$; negative symptoms, $d = 0.38$, $p < 0.05$) than participants from HIC, but a significantly lower score on avolition ($d = -0.42$, $p < 0.05$).

Furthermore, the CAPE with the adjusted factorial structure (MI model) produced different mean scores within the country groups than the non-adjusted version of the CAPE (non-MI model). As can be seen in Table 4, the latent mean differences for paranoia, magical thinking and avolition between LAMIC and HIC changed statistical significance after MI adjustment. For example, paranoia in the non-MI model was found to be significantly higher in LAMIC than HIC ($d = 38.26$, $p < 0.05$), but this difference was no longer significant after MI adjustment ($d = 0.03$, $p = 0.89$). In contrast, avolition was found to not differ in LAMIC and HIC ($d = 0.03$, $p = 0.64$), but after MI adjustment avolition was significantly lower in LAMIC than HIC ($d = -0.42$, $p < 0.05$). Despite numerical differences in *Cohen's d* estimates between the non-MI and MI adjusted models, we found the effect size category to remain the same for the remaining factors of the CAPE.

Discussion

We examined whether measurement of PEs with the CAPE is invariant between LAMIC and HIC in order to understand whether measurement bias explains previously found differences in levels of PEs across country income groups (McGrath et al., 2015). We found configural and metric MI which means that the structure and understanding of PEs in the CAPE are comparable between LAMIC and HIC (e.g. positive symptoms factor, negative symptoms factor and depression factor are separate constructs, hallucinations factor are a part of the positive symptoms factor, hearing voices item is a part of the hallucinations factor). However, we did not find full scalar MI of the CAPE. Here, we found that the grandiosity factor, the hallucinations factor, and the items being influenced by devices, thought echo, and the lack of hygiene was noninvariant, meaning that their differing levels across LAMIC and HIC do not indicate true latent score differences between the groups. Moreover, we did not find residual MI, which means that differences between LAMIC and HIC in observed PE scores are partly explicable by measurement bias.

We found positive symptoms to be most affected by MI adjustment. This aligns with findings from studies testing MI of other scales such as the SPQ-B (Fonseca-Pedrero et al., 2018), the Wisconsin Schizotypy Scale (Cicero et al., 2019*b*), and the PSQ (Heuvelman et al., 2018) that also identified problems in measuring positive symptoms invariantly across culture groups. Specifically, we found hallucinations, grandiosity, and paranoia to be most affected by MI adjustment. The lack of MI regarding the hallucinations and grandiosity factor of the CAPE has also been shown in a student sample study (Vermeiden et al., 2019).

Table 3. Measurement invariance of the CAPE between LAMIC ($n = 2472$) and HIC ($n = 4669$, $N = 7141$)

	Robust χ^2									
Model	χ^2	*df*	*p*	CFI	Δ CFI	RMSEA	Δ RMSEA	SRMR	Δ SRMR	Decision
Configural invariance model	21 050.64	1618	<0.01	0.969		0.039		0.050		
Metric invariance model	20 482.70	1657	<0.01	0.961	0.008	0.043	0.004	0.054	0.004	Accept
Scalar invariance model	23 806.31	1688	<0.01	0.954	0.015	0.047	0.008	0.056	0.006	Reject
Partial scalar invariance model	20 578.19	1681	<0.01	0.962	0.007	0.042	0.003	0.053	0.003	Accept
Residual invariance model	27 140.49	1730	<0.01	0.947	0.022	0.050	0.010	0.064	0.014	Reject
Partial residual invariance model	24 281.61	1723	<0.01	0.955	0.014	0.046	0.007	0.061	0.011	Reject

Note. LAMIC = Low-and-Middle Income-Countries ($n = 2472$). HIC = High Income-Countries ($n = 4669$). CFI = Robust Comparative Fit Index. RMSEA = Robust Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. Metric invariance model was compared with configural invariance model. Scalar and partial scalar invariance model was compared with metric invariance model. Residual and partial residual invariance model was compared with scalar invariance model.

Table 4. Latent mean comparison of the CAPE factors between LAMIC ($n = 2472$) and HIC ($n = 4669$, $N = 7141$)

Factors	$d_{uncorrected}$	$d_{corrected}$
Positive symptoms	1.15*	0.83*
Bizarre experiences	0.47*	0.54*
Hallucinations	0.03	–
Paranoia	38.26*	0.03
Grandiosity	0.93*	–
Magical thinking	−0.03	0.33*
Negative symptoms	0.42*	0.38*
Social withdrawal	0.06*	0.15*
Blunted affect	0.74*	0.75*
Avolition	0.03	−0.42*
Depression	0.18*	0.18*

Note. * $p < 0.05$. $d_{uncorrected}$ = Latent mean differences across comparison group based on non-measurement-invariant model. $d_{corrected}$ = Latent mean differences across comparison group based on measurement-invariant model LAMIC = Low-and-Middle Income-Countries ($n = 2472$). HIC = High Income-Countries ($n = 4574$). Positive values indicate LAMIC>HIC. Negative values indicate LAMIC<HIC. There are no scores for hallucinations and grandiosity factors in the measurement-invariant model because they are non-invariant.

Both Vermeiden et al. (2019) and our study found the paranoia factor of the CAPE to have MI across culture groups. However, both Heuvelman et al. (2018) and our study found paranoia factor differences between culture group changes after MI adjustment. They found the paranoia questions in PSQ overestimate the levels of paranoia in the Pakistani ethnic group in comparison to White British in the UK (Heuvelman et al., 2018), while we found the paranoia factor of the CAPE to be overestimated in LAMIC in comparison to HIC. Thus, the measurement bias we found in relation to paranoia may not be a specific problem of the CAPE. Rather, it might indicate a general difficulty of constructing an invariant measure of paranoia. This difficulty may stem from the fact that a measure of paranoia has to take objective levels of threat into account, which are likely to differ between geographical locations. For example, people in some countries may objectively face more threat (e.g. due to minority sexual status), which would be reflected in their responses to the CAPE paranoia items (e.g. 'Do you ever feel as if you are being persecuted in some way?', 'Do you ever feel that people look at you oddly because of your appearance?').

In contrast, MI adjustment only had a minimal effect on the depression and negative symptoms factor of the CAPE. This is consistent with previous studies showing complete MI of depression and negative symptoms measures across different cultural groups. For example, the Patient Health Questionnaire-9 has been found to be MI in adult Dutch and Surinam Dutch patients in the Netherlands (Baas et al., 2011) and in community samples of English and Spanish speaking Hispanic American women in the USA (Merz, Malcarne, Roesch, Riley, & Sadler, 2011). The Center for Epidemiological Studies Depression scale has also been found to be MI on community samples of Chinese and Dutch older adults (Zhang et al., 2011). The Brief Negative Symptoms Scale has been found to have residual MI in Chinese, English, German, Italian and Spanish patients (Ahmed et al., 2019). Thus, MI seems to be less of an issue in assessing depression and negative symptoms across countries and ethnic backgrounds than in assessing positive symptoms.

Even after MI corrections, LAMIC continued to show higher levels of some positive symptoms (bizarre experiences and magical thinking) and negative symptoms (social withdrawal and blunted affect). This speaks against the assumption that the higher levels of PEs are entirely explicable by MI. Thus, higher PEs in LAMIC than HIC could reflect cultural differences (World Health Organization, 2011), a poorer mental health status (McGrath et al., 2016), or both. The interpretation of poorer mental health in LAMIC is indirectly supported by a Lancet Commission Report (Patel et al., 2018), which points to the problematic discrepancy between the high prevalence of social determinants of poor mental health (e.g. higher levels of poverty, higher number of loss, trauma, and displacement due to war and disasters) and the low availability of mental health services in LAMIC compared to HIC (e.g. median number of mental health professionals 10.1 *v.* 50.8 per 100 000).

Nevertheless, the paradox remains that despite the somewhat higher levels of PEs in the population, the rates of clinically relevant and diagnosable psychotic disorders are lower in LAMIC than in HIC (McGrath et al., 2008). A putative explanation for

https://doi.org/10.1017/S0033291720003323 Published online by Cambridge University Press

this paradox can be derived from the theory of extended psychosis phenotype (van Os & Reininghaus, 2016), which states that an accumulation of PEs is an essential feature of transition. As there is an indication of lower PE-related distress in LAMIC compared to HIC (Wüsten et al., 2018) and non-distressing PEs have been found to be less tightly interrelated with other PEs in network models (Murphy, McBride, Fried, & Shevlin, 2018), lower distress may be a protective factor in this regard. In addition, the low rates of psychotic disorders in LAMIC (*v.* HIC) may partly be due to underdiagnosis due to restricted access to medical facilities and the presence of other ways of seeking help (e.g. traditional healers).

Strengths and limitations

A strength of this study is that the community samples were recruited from different regions and cultures. Countries from the continents of Asia, Africa, North America, South America, and Europe were represented in the study. This allows us to interpret the findings as generalizable with some confidence. Furthermore, a self-report assessment, in comparison to interviews, has the advantage of being free of interviewer bias and rating inconsistencies across countries.

A limitation of the study is the heterogeneous context of the study, the participant recruitment process, and the assessment type (online *v.* paper-and-pencil). For example, the sample from Ghana was recruited entirely online, whereas the sample from the Netherlands/Belgium was recruited entirely via a paper-and-pencil procedure. The assessment type differences between the country income group may have introduced selection bias, which may be a source of non-MI. Besides differences in assessment type proportion between the country income groups, the significant differences in sex and age between the country income groups may also be a source of non-MI. We cannot rule out that some of the problems in MI are attributable to these factors rather than country income group. Similarly, MI problems could also arise from differences in the language version of the measure rather than from country income group. These limitations should be addressed in future studies by (a) taking care to balance samples in terms of potentially relevant factors and (b) by conducting MI analyses on further variables, such as assessment type, countries, language, age, gender, and socioeconomic status.

Conclusion

We found full configural and metric MI of the CAPE, indicating that there is a culturally universal agreement on the structure and conceptualization of PEs across LAMIC and HIC. However, we only succeeded in obtaining partial scalar MI, indicating that we can conduct latent, but not observed sum-score mean comparisons without risk of bias. After MI adjustment, we found several PEs from the positive symptom domain to be higher in LAMIC than HIC. Thus, the paradox remains that the prevalence of psychotic disorders is lower whereas subclinical PEs are higher in LAMIC than in HIC.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291720003323.

Acknowledgements. The authors are grateful to Steffi Hartanto who assisted with the data analysis. Edo S. Jaya was supported by the Universitas Indonesia's internal research grant ('Hibah Publikasi Artikel di Jurnal Internasional Kuartil Q1 dan Q2 Tahun Anggaran 2019', Grant No: NKB-0281/UN2.R3.1/HKP.05.00/2019). Eduardo Fonseca Pedrero was supported by the Spanish Ministry of Science and Innovation (MICINN) and by the Instituto Carlos III, Center for Biomedical Research in the Mental Health Network (CIBERSAM) and by the 'Convocatoria 2015 de Ayudas Fundación BBVA a Investigadores y Creadores Culturales' and the 'Ayudas Fundación BBVA a Equipos de Investigación Científica 2017'.

Conflict of interest. None.

References

Ahmed, A. O., Kirkpatrick, B., Galderisi, S., Mucci, A., Rossi, A., Bertolino, A., … Strauss, G. P. (2019). Cross-cultural validation of the 5-factor structure of negative symptoms in schizophrenia. *Schizophrenia Bulletin*, *45*, 305–314.

Al-Issa, I. (1995). The illusion of reality or the reality of illusion: Hallucinations and culture. *The British Journal of Psychiatry*, *166*, 368–373.

Baas, K. D., Cramer, A. O. J., Koeter, M. W. J., van de Lisdonk, E. H., van Weert, H. C., & Schene, A. H. (2011). Measurement invariance with respect to ethnicity of the patient health questionnaire-9 (PHQ-9). *Journal of Affective Disorders*, *129*, 229–235.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588.

Brenner, K., Schmitz, N., Pawliuk, N., Fathalli, F., Joober, R., Ciampi, A., & King, S. (2007). Validation of the English and French versions of the community assessment of psychic experiences (CAPE) with a Montreal community sample. *Schizophrenia Research*, *95*, 86–95.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 464–504.

Cicero, D. C., Krieg, A., & Martin, E. A. (2019a). Measurement invariance of the prodromal questionnaire–brief among white, Asian, Hispanic, and multiracial populations. *Assessment*, *26*, 294–304.

Cicero, D. C., Martin, E. A., & Krieg, A. (2019b). Differential item functioning of the full and brief wisconsin schizotypy scales in Asian, White, Hispanic, and multiethnic samples and between sexes. *Assessment*, *26*, 1001–1013.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New Jersey: L. Erlbaum Associates.

DeVylder, J. E., & Koyanagi, A. (2018). Evaluating the clinical relevance of psychotic experiences in low- and middle-income countries. *Schizophrenia Bulletin*, *44*, 1167–1169.

Fonseca-Pedrero, E., Ortuño-Sierra, J., Lucas-Molina, B., Debbané, M., Chan, R. C. K., Cicero, D. C., … Voracek, M. (2018). Brief assessment of schizotypal traits: A multinational study. *Schizophrenia Research*, *197*, 182–191.

Fonseca-Pedrero, E., Paino, M., Lemos-Giráldez, S., & Muñiz, J. (2012). Validation of the community assessment psychic experiences −42 (CAPE-42) in Spanish college students and patients with psychosis. *Actas Españolas de Psiquiatría*, *40*, 169–176.

Heuvelman, H., Nazroo, J., & Rai, D. (2018). Investigating ethnic variations in reporting of psychotic symptoms: A multiple-group confirmatory factor analysis of the psychosis screening questionnaire. *Psychological Medicine*, *48*, 2757–2765.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55.

Jablensky, A. (2000). Epidemiology of schizophrenia: The global burden of disease and disability. *European Archives of Psychiatry and Clinical Neuroscience*, *250*, 274–285.

Jablensky, A., Sartorius, N., Ernberg, G., Anker, M., Korten, A., Cooper, J. E., … Bertelsen, A. (1992). Schizophrenia: Manifestations, incidence and course in different cultures. A World Health Organization ten-country study. *Psychological Medicine Monograph*, *20*, 1–97.

Jaya, E. S. (2017). Confirmatory factor analysis of the Indonesian version of community assessment of psychic experiences. *Makara Human Behavior Studies in Asia*, *9*, 1–12.

Korver, N., Quee, P. J., Boos, H. B. M., Simons, C. J. P., & de Haan, L., & GROUP investigators (2012). Genetic risk and outcome of psychosis (GROUP), a multi site longitudinal cohort study focused on gene–

environment interaction: Objectives, sample characteristics, recruitment and assessment methods. *International Journal of Methods in Psychiatric Research*, *21*, 205–221.
Luhrmann, T. M., Padmavati, R., Tharoor, H., & Osei, A. (2015). Differences in voice-hearing experiences of people with psychosis in the USA, India and Ghana: Interview-based study. *The British Journal of Psychiatry: The Journal of Mental Science*, *206*, 41–44.
Mark, W., & Toulopoulou, T. (2016). Psychometric properties of 'community assessment of psychic experiences': Review and meta-analyses. *Schizophrenia Bulletin*, *42*, 34–44.
McGrath, J. J., Saha, S., Al-Hamzawi, A., Alonso, J., Bromet, E. J., Bruffaerts, R., … Kessler RC (2015) Psychotic experiences in the general population: A cross-national analysis based on 31 261 respondents from 18 countries. *JAMA Psychiatry*, *72*, 697–705.
McGrath, J. J., Saha, S., Al-Hamzawi, A., Andrade, L., Benjet, C., Bromet, E. J., … Kessler, R. C. (2016). The bi-directional associations between psychotic experiences and DSM-IV mental disorders. *The American Journal of Psychiatry*, *173*, 997–1006.
McGrath, J. J., Saha, S., Chant, D., & Welham, J. (2008). Schizophrenia: A concise overview of incidence, prevalence, and mortality. *Epidemiologic Reviews*, *30*, 67–76.
Merz, E. L., Malcarne, V. L., Roesch, S. C., Riley, N., & Sadler, G. R. (2011). A multigroup confirmatory factor analysis of the patient health questionnaire-9 among English- and Spanish-speaking Latinas. *Cultural Diversity & Ethnic Minority Psychology*, *17*, 309–316.
Murphy, J., McBride, O., Fried, E., & Shevlin, M. (2018). Distress, impairment and the extended psychosis phenotype: A network analysis of psychotic experiences in a US general population sample. *Schizophrenia Bulletin*, *44*, 768–777.
Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., … Jü, U. (2018). The lancet commission on global mental health and sustainable development. *The Lancet*, *392*, 1553–1598.
Peters, E., Ward, T., Jackson, M., Morgan, C., Charalambides, M., McGuire, P., … Garety, P. A. (2016). Clinical, socio-demographic and psychological characteristics in individuals with persistent psychotic experiences with and without a 'need for care'. *World Psychiatry*, *15*, 41–52.
Pignon, B., Peyre, H., Ferchiou, A., van Os, J., Rutten, B. P. F., Murray, R. M., … Szöke, A., & EU-GEI WP2 Group Author (2018) Assessing cross-national invariance of the Community Assessment of Psychic Experiences (CAPE). *Psychological Medicine*, *49*, 2600–2607.
Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90.
Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.
Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507–514.
Schlier, B., Jaya, E. S., Moritz, S., & Lincoln, T. M. (2015). The community assessment of psychic experiences measures nine clusters of psychosis-like experiences: A validation of the German version of the CAPE. *Schizophrenia Research*, *169*, 274–279.
Siddi, S., Ochoa, S., Laroi, F., Cella, M., Raballo, A., Saldivia, S., … Preti, A. (2019). A cross-national investigation of hallucination-like experiences in 10 countries: The E-CLECTIC study. *Schizophrenia Bulletin*, *45*, S43–S55.
Stefanis, N. C., Hanssen, M., Smirnis, N. K., Avramopoulos, D. A., Evdokimidis, I. K., Stefanis, C. N., … van Os, J. (2002). Evidence that three dimensions of psychosis have a distribution in the general population. *Psychological Medicine*, *32*, 347–358.
van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*, 486–492.
van Os, J. (2016). Measures of psychosis proneness and genetic risk for schizophrenia. *JAMA Psychiatry*, *73*, 638.
van Os, J., & Reininghaus, U. (2016). Psychosis as a transdiagnostic and extended phenotype in the general population. *World Psychiatry*, *15*, 118–124.
Verdoux, H., Sorbara, F., Gindre, C., Swendsen, J. D., & van Os, J. (2003). Cannabis use and dimensions of psychosis in a nonclinical population of female subjects. *Schizophrenia Research*, *59*, 77–84.
Vermeiden, M., Janssens, M., Thewissen, V., Akinsola, E., Peeters, S., Reijnders, J., … Lataster, J. (2019). Cultural differences in positive psychotic experiences assessed with the Community Assessment of Psychic Experiences-42 (CAPE-42): A comparison of student populations in the Netherlands, Nigeria and Norway. *BMC Psychiatry*, *19*, 244.
World Bank (2016). World Bank Country and Lending Groups. Available at: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.
World Health Organization (2011). *Mental health atlas 2011*. Italy: WHO Press.
Wüsten, C., Schlier, B., Jaya, E. S., Genetic Risk and Outcome of Psychosis (GROUP) Investigators, Fonseca-Pedrero, E., Peters, E., Verdoux, H., … Lincoln, T. M. (2018) Psychotic experiences and related distress: A cross-national comparison and network analysis based on 7141 participants from 13 countries. *Schizophrenia Bulletin*, *44*, 1185–1194.
Yoon, M., & Lai, M. H. C. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 201–213.
Zhang, B., Fokkema, M., Cuijpers, P., Li, J., Smits, N., & Beekman, A. (2011). Measurement invariance of the center for epidemiological studies depression scale (CES-D) among Chinese and Dutch elderly. *BMC Medical Research Methodology*, *11*, 74.
Ziermans, T. B. (2013). Working memory capacity and psychotic-like experiences in a general population sample of adolescents and young adults. *Frontiers in Psychiatry*, *4*, 161. doi: 10.3389/fpsyt.2013.00161.