International Journal of Psychology

International Journal of Psychology

EMPIRICAL ARTICLE

Wellbeing Dynamics Across Cultures During the Pandemic: A Five-Country Exploration

Larissa Hartle[1,2] | Tiago Bortolini[1,2] | Johannes Karl[3] | Boris Sokolov[4] | Ronald Fischer[2,5]

[1]D'Or Institute for Research and Education, Rio de Janeiro, Brazil | [2]Pioneer Science Initiative, Rio de Janeiro, Brazil | [3]Dublin City University, Dublin, Ireland | [4]HSE University, Moskva, Russia | [5]D'Or Institute for Research and Education, São Paulo, Brazil

Correspondence: Larissa Hartle (larissa.hartle@idor.org)

Received: 19 May 2024 | **Revised:** 17 December 2024 | **Accepted:** 24 December 2024

Funding: This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (process: 21/08774–1) supported work by Ronald Fischer. Boris Sokolov's work was supported by the HSE University Basic Research Program. Larissa Hartle's work was supported by the IDOR/Pioneer Science Initiative.

Keywords: cross-cultural | factorial structure | invariance | longitudinal | wellbeing

ABSTRACT

Wellbeing levels have been a global concern during the COVID-19 pandemic, but there is a lack of attention to invariance questions that allow a robust examination of wellbeing dynamics across cultures. Questions of temporal stability that are crucial for examining the impact of the pandemic on wellbeing have received even less attention. Some studies suggested that measures may not be stable after the onset of the pandemic. We examine invariance parameters, the factorial structure and variability of wellbeing variables (life satisfaction, pandemic worries, anxiety and depression screenings) across five different cultural contexts from 2020 to 2022 ($N = 4387$, total observations = 13,161). A three-factor model separating life satisfaction, worry and distress performed best in terms of model fit and parsimony. We observed scalar invariance across times and identified little variability of wellbeing measures during the pandemic, suggesting that wellbeing levels remained stable during the pandemic in each of the countries sampled. In contrast, we only identified metric invariance across countries at each time point, and found a weakening of correlations between life satisfaction and a depressive/anxious symptoms scale in lower income countries. We discuss implications of our findings for discussions of wellbeing dynamics.

1 | Introduction

The COVID-19 pandemic brought with it profound upsets to people's lifestyles compounded with fluctuating and distressing rates of contagion and mortality. Perhaps unsurprisingly, studies have shown elevated levels of stress and lower levels of wellbeing (Manão et al. 2022; Zacher and Rudolph 2021). During the pandemic, a unique opportunity to investigate an important question emerged: do psychological scales change their meaning across contexts or across time, especially if individuals are facing acute changes in their daily routines and habits? Research early in the pandemic suggested that individuals started interpreting personality scale items differently after the onset of the pandemic (Sutin et al. 2020), and well-established wellbeing measures showed only configural invariance across cultures during the same period (Yap et al. 2021).

A recent study highlighted that part of the replication crisis within psychology may be due to the lack of attention to invariance of psychological measures (Maassen et al. 2023). So far, questions of comparability of measures have been primarily addressed in a culture-comparative context, and relatively less work has focused on constructs and their empirical measures across time. Maassen et al. (2023) highlight that inquiries of

invariance are notoriously underreported in psychological science in general, and yet the lack to demonstrate measurement equivalence across multiple measurement occasions has substantive implications for the interpretation of any differences.

Briefly, the invariance paradigm examines the similarity or differences of the relationship between an item and the implied latent variable across studied groups or across time within the same group (Leitgöb et al. 2023). Typically, three levels can be distinguished in measurement invariance: configural invariance, metric invariance and scalar invariance. The first level relates to item—instrument coherence, the second level to the discrimination capacity of the item—or the factor loadings in statistical terms, and the third level to the item difficulty, which is the intercept or thresholds in statistical terms. If all of these are similar across groups, full score invariance has been achieved. Full score invariance is important because it implies that means of a latent variable can be directly compared. Testing measurement invariance can bring more certainty to comparative research because it rules out other sources of difference other than the main measured variable itself, that is, if the measure has the same properties across groups, whether the groups are time points or countries.

The identification of measurement invariance at any particular level has also important theoretical implications for the overall construct space. If we were to find scalar invariance across time, it would imply that the measured construct remained stable over time even in a period of acute crisis. If we were to find metric equivalence, it would imply that response to measurement instruments is only partially driven by the intended wellbeing concept and that other psychological processes may play a role in affecting responses to individual items. If we were to find only configural invariance, it would imply that expression of the specific wellbeing construct itself changed its meaning during the pandemic.

In our current study, we examine invariance parameters of a domain specific measure of life satisfaction, pandemic worries and a brief mental health screening tool over three time points and across five different cultural contexts during the recent COVID-19 pandemic. Reporting nearly representative data from five distinct cultural groups we provide new insights into the structure and stability of wellbeing measures during a period when psychological measures may have been affected by the great uncertainties and disruptions. The resulting patterns provide greater confidence about mental health and wellbeing measurement in populations across the world. We next briefly review cross-cultural invariance investigations followed by temporal invariance investigations.

1.1 | Cross-Cultural Invariance of Wellbeing Measures

A number of previous studies have examined cross-cultural invariance parameters within a broad wellbeing domain at a single time point. A recent review (Emerson, Guhn, and Gadermann 2017) identified 11 studies investigating measurement invariance of the widely used Satisfaction with life scale (Diener et al. 1985), which captures a general sense of satisfaction with life. In line with broader cross-cultural invariance investigations (Rutkowski and Svetina 2014), only one study supported scalar measurement invariance when comparing data from university students in the US and Russia, but only configural invariance was found when comparing adults. The remainder studies from the review reported metric invariance for four studies, configural invariance for another four studies, and no configural invariance for two (Emerson, Guhn, and Gadermann 2017). Focusing specifically on studies conducted during the pandemic, the same satisfaction with life scale reached only metric invariance across samples of Chinese and Euro-Canadian students during the early stages of the COVID-19 pandemic (Yap et al. 2021). In contrast, Caycho-Rodríguez et al. (2023) reported scalar invariance across convenience samples of mainly university-educated individuals from 12 Latin American countries (mean sample size $N = 432$) using the WHO-5 wellbeing questionnaire in the first few months of the pandemic. All these samples came from the same continent and all but one used the same language and belonged to the same cultural cluster identified by Inglehart and Welzel (2005). Therefore, the larger or more culturally different the sample, the harder it seems to be to reach scalar measurement invariance (Rutkowski and Svetina 2014). The studies using life satisfaction conducted during the pandemic also suggest that the pandemic context may impact the measurement of wellbeing.

Context-specific wellbeing measures developed and tested during the pandemic showed similar problems. The Fear of COVID-19 Scale reached partial metric invariance across 48 countries only when using more lenient thresholds (Sawicki et al. 2022). Other patterns reported in the literature suggested that cultural distance matters, because the same scale reached metric invariance across 3 Portuguese-speaking countries in a separate study (Giolo et al. 2022). These studies raise concerns about the possibility to compare wellbeing scores across cultures during the pandemic, which may limit our ability to discuss the differential impact of the pandemic on mental wellbeing as well as recovery of mental wellbeing as the pandemic recedes. Thus, testing for invariance is of special importance in order to draw meaningful insights into the differential impact and trajectory of the pandemic across cultures.

1.2 | Longitudinal Invariance of Wellbeing Measures

Shifting the focus to measurement invariance over time, questions of temporal stability have received even less attention. One issue is the relative absence of longitudinal studies in psychology more generally. The few studies that have directly examined these dynamics over time show a similarly complex pattern in line with the cross-national invariance literature. One study found scalar invariance across four time points over a two-year period for the Habitual Subjective Wellbeing Scale (HSWBS), but not even configural invariance was found for the more complex Multidimensional Student's Life Satisfaction Scale (MSLSS) measuring satisfaction with family and peers (Steinmayr, Paschke, and Wirthwein 2022). Another study (Zacher and Rudolph 2021) reported metric but not scalar invariance across five time points over 3 years for factors measuring positive and negative affect. Similarly, Lanz et al. (2023) only identified metric invariance across three time points in scales measuring stress or psychological wellbeing.

These studies suggest that measures may change their properties across time. This is of particular importance when examining wellbeing during an event such as the pandemic, which fundamentally disrupted daily life and may have changed how individuals interpret standard wellbeing measures, often developed outside such extraordinary contexts. Any changes in the measurement will have implications for claims about mental health during and in the recovery process of a pandemic. If the instruments are not measuring the same construct over time, it may not be possible to track how the pandemic impacted wellbeing within the same populations. We therefore explicitly ask: Do wellbeing concepts remain stable over a dynamic period such as the COVID-19 pandemic? Would any structural differences vary across different domains of wellbeing (e.g., more clinical concerns of negative wellbeing which may be more impacted by the pandemic versus more positively phrased satisfaction with life measures covering different life domains that may have been impact by the pandemic)?

1.3 | Wellbeing Variables

We examine these patterns using three different measures: (1) domain specific life satisfaction, (2) worries and (3) anxiety and depressive symptoms. From a cognitive perspective, wellbeing is the presence of positive thoughts about the self and one's life and the absence of negative thought patterns. One commonly studied cognitive component is life satisfaction, which is a person's cognitive evaluation of their own life (Diener 1984). Life satisfaction can be measured using both global or domain specific evaluations (covering satisfaction with specific life domains), with domain specific representations possibly influencing global life satisfaction (bottom-up) or global perceptions guiding domain specific evaluations (top-down) (Diener 1984; Fischer and Van de Vliert 2011; Lance, Mallard, and Michalos 1995; Leonardi, Spazzafumo, and Marcellini 2005; Locke 2004; Nakazato, Schimmack, and Oishi 2011). Much of the cross-cultural invariance work has focused on global measures and less is known about domain specific evaluations. Domain specific measures might be particularly interesting to investigate during a period of upheaval, such as the pandemic, because individuals need to re-evaluate their satisfaction with specific domains given the changed life conditions. Therefore, we focus on a domain specific measure of life satisfaction that captures satisfaction with different domains that may have been differentially affected by the pandemic.

Complementary to this general evaluation of life domains, worries, are negatively valenced cognitive representations of specific threats (Schwartz, Sagiv, and Boehnke 2000). Important worries during the pandemic were related to one's health and safety, the safety of close others as well as concerns about the state of the one's society, including existential worries about the economic implications of the pandemic (Lanz et al. 2023; Manão et al. 2022; Mónaco et al. 2022). Finally, when worries occur more chronically and interfere with daily life activities, wellbeing deteriorates and clinical symptoms of anxiety or closely related symptoms of depression may emerge (Beuke, Fischer, and McDowall 2003). Together, the three wellbeing variables allow approaching wellbeing from multiple perspectives. Our investigation therefore also allows for a more data driven approach to their empirical stability and empirical distinctiveness.

1.4 | The Current Study

We use population level data from high-quality opt-in panels from 5 countries to test both cross-cultural invariance as well as temporal invariance across the three time points during the COVID-19 pandemic. Specifically, we use samples from 5 different cultural regions: Catholic Europe (Austria), Protestant Europe (Germany), Orthodox Europe (Russia), Anglo (UK) and Latin American (Brazil) cultural clusters as identified in data from the World Values Survey as the longest running and largest multinational comparative survey of human attitudes and values (Inglehart and Welzel 2005). The culturally diverse characteristics including samples from three continents together with the near-representative nature of our samples offers significant advantages over testing wellbeing structures in samples of university students, university setting or convenience sampling. Focusing on the stability across time and samples provides important information on the stability versus variability of wellbeing as a broad psychological construct during a period of profound societal upheaval. Depending on the results of the invariance tests, we also aim to compare wellbeing across countries and across the pandemic as well as the relationship between the three wellbeing variables over time to provide further insights into how different wellbeing measures relate to each other over time.

2 | Material and Methods

2.1 | Data Collection

Data were gathered for the 'Values in Crisis' project (Aschauer et al. 2021; Welzel et al. 2023) through online panels targeting representative data. We included countries with three data points and the specific data collection points varied somewhat by country (see Figure S1). The first phase of data collection began for all countries in the first 3 months of the pandemic in 2020, the second data collection phase took place during the first semester of 2021, and the final wave was collected after vaccination campaigns had already started in all countries in our sample (December 2021 – July 2022). In Austria data were collected by the Marketagent.com online reSEARCH GmbH. In Brazil, nationally representative data were collected online by the Sivis Institute. In Germany and the United Kingdom, data collection were carried out by bilendi GmbH. In Russia, data were collected via an online opt-in panel operated by the polling organisation Online Marketing Intelligence (OMI). Due to dropout rates and the use of data of subjects who responded to all three waves, the data only remained approximately representative for certain characteristics in some countries. For demographic details for all samples, see Table 1.

2.2 | Wellbeing Measures

To measure domain specific life satisfaction, we used 5 questions with answers recorded on a scale from 1 (completely unsatisfied) to 10 (completely satisfied). Questions were preceded by 'All things considered, how satisfied are you these days with your . . . ' and followed by (1) health condition, (2) financial situation, (3) social relations, (4) work-life balance and (5) life as a whole. This scale was developed by the Values in Crisis Project (Welzel et al. 2023) to cover different life domains that may have been

TABLE 1 | Sample description at Wave 1.

Variable	Austria $N = 747$	Brazil $N = 1301$	Germany $N = 878$	Russia $N = 694$	UK $N = 767$
Mean age	53.15 (15.00)	47.41 (14.71)	50.06 (13.42)	50.38 (13.16)	52.77 (16.02)
Age range	[18–81]	[18–91]	[18–77]	[19–84]	[18–88]
Sex					
Male	48.45%	50.73%	53.18%	53.45%	55.54%
Female	51.54%	49.26%	46.81%	46.54%	44.45%

differentially affected by the disruptions caused by the pandemic. This scale has been shown to have a Cronbach's alpha of 0.84 in the first wave and 0.85 in the second wave of the study (Kerstis et al. 2021).

To measure worry, we measured COVID infection and economic recession worries with one item each: 'How afraid are you that you or your loved ones get sick and suffer severely from the Corona virus?'; 'How afraid are you that you or your loved ones will suffer from an economic recession following the Corona crisis?'. Participants responded on a 5-points Likert scale (from 'not afraid at all' to 'very afraid'). These two items were developed for the Values in Crisis project, building on earlier work by Schwartz, Sagiv, and Boehnke (2000).

Assessing depression and anxiety symptoms in general population samples, the PHQ-4 (Löwe et al. 2010) is a standardised screening instrument for generalised anxiety and depression. Items were responded to on a 4-points scale indicating the extent to which people have been bothered by specific problems in the last 2 weeks ranging from 'not at all' to 'practically everyday'. In general, the most frequent factor structure used is two factors with two items per factor (Caro-Fuentes and Sanabria-Mazo 2024), although Wicke et al. (2022) calls the separation of these factors into question given the high correlation of the two factors. A systematic review found internal consistency was adequate for the four-item PHQ-4 scale (α from 0.72 to 0.88), as well as for the two items measuring depression (PHQ-2; Cronbach's alphas from 0.65 to 0.81) and anxiety (GAD-2; Cronbach's alphas from 0.74 to 0.84) (Caro-Fuentes and Sanabria-Mazo 2024).

Cronbach's alphas and McDonald's Omegas for this study are reported in Table S1 for all waves and samples. All values and wellbeing scores were converted to the proportion of maximum possible scores (POMP; Cohen et al. 1999). Resulting values range from 0 to 1, with higher scores indicating higher wellbeing (see Table S2 for *N*, mean scores and standard deviations).

2.3 | Data Analysis

To investigate the invariance of wellbeing measures, it was first necessary to examine the underlying structure of the wellbeing questionnaire across different cultures and waves. We first conducted a confirmatory factor analysis (CFA) with all respondents using the *cfa* function from the *lavaan* package in R (R Core Team 2022), with Maximum Likelihood as the estimation method. We tested solutions ranging from 1 to 4 factors to identify the optimal number of factors that best capture the underlying constructs in the questionnaire responses. The analysis was performed on all items (5 life satisfaction items, 2 worry items and 4 items from the Patient Health Questionnaire — PHQ-4) across all waves and countries. The model with one factor had the 5 life satisfaction items, 4 PHQ-4 items and 2 worry items loading together on a single factor. The model with two factors had life satisfaction items loading on the first factor, and a factor combining worry and PHQ-4 items, thereby forming separate positive and negative wellbeing factors. The three-factor model had each wellbeing construct loading on a separate factor. Finally, the four-factor model separated the 'depression' and 'anxiety' factors from PHQ-4. After identifying the optimal number of factors, we subsequently computed fit indices for each individual wave and country.

We use as fit indicators the chi-squared statistic, the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI) and the Standardised Root Mean Square Residual (SRMR). While there are no universally agreed-upon cut-off points, generally, RMSEA values below 0.08, CFI and TLI values above 0.90 or 0.95, and SRMR values below 0.08 are indicative of acceptable fit (Hu and Bentler 1999), but interpretation should consider the specific context and purpose of the analysis.

To assess invariance of the optimal model across time and countries, we used the *equival* function from the *ccpsyc* package, following the methodology outlined by Fischer and Karl (2019). The function calculates fit indices for different models, one with no constraints (configural invariance), another with loading constraints (metric invariance), and a third with both loading and intercept constraints (scalar invariance). Differences larger than 0.02 observed in the McDonald Non-Centrality Index and larger than 0.01 in the CFI across the models signify invariance is rejected. First, for temporal invariance, we ran the *equival* function adopting the model derived from prior confirmatory factor analyses. We computed the configural, metric and scalar invariance, with the 'wave' variable as the group parameter separately for each cultural sample. Similarly, for cultural invariance assessment, we executed a CFA for each wave with the 'country' variable as the group parameter. This two-fold approach allowed us to comprehensively evaluate the stability of the measurement across time and cultures.

As an additional analysis, we addressed the variability of wellbeing across countries and over time and the relationship between

the wellbeing variables. First, we utilised the *aov* function in R to compare mean wellbeing scores across different time points within each country. Second, we correlated wellbeing measures considering each country and each wave. At last, we compared these correlations across cultures and time. To do so, we used r-to-z transformations and Fisher's *Z* tests using the *cocor* function. For all tests, we applied Bonferroni corrections to adjust for multiple comparisons.

3 | Results

3.1 | Structure

We first decided on the optimal number of factors to explore further. Examining the fit indices and the overall patterns, we decided that the three-factor model was the more suitable model. The fit indices for the three-factor solution were above commonly recommended thresholds, suggesting that this structure showed appropriate fit and it was more parsimonious. Although separate factors for the PHQ-4 (differentiating anxiety and depression) showed somewhat better fit, we are cautious about using this structure for three main reasons. First, the two components had strong zero-correlations in our data ($r > 0.7$). Second, the instrument is intended as an overall screening tool and the separate factors may not have sufficient breadth to clearly measure anxiety and depressive symptoms in a general population. Finally, and importantly from a practical perspective, the two components only had 2 items per scale, which may result in identification issues. Invariance results did not change comparing the three-factor and four-factor models. Fit indices for the three-factor model were good when considered all responses collectively or when we computed fit indices for each wave and country independently. Table 2 reports the fit indices chi-squared, CFI, Non-Normed Fit Index (NNFI), Root Mean Square Error of Approximation (RMSEA) and Standardised Root Mean Residual (SRMR) for each model, and Table 3 reports the same fit indices for the three-factor model considering each wave and country. Table S3 shows fit indices for the four-factor models. Examining the covariance between the individual wellbeing constructs, the covariance between PHQ-4 and Life satisfaction was 0.445, PHQ-4 and Worry was 0.431, and life satisfaction and worry 0.294. These patterns suggest reasonable convergence between the measures.

3.2 | Invariance

We then ran separate invariance tests over time and over countries. chi-squared, CFI, RMSEA, Gamma, MNCI and differences between constrained models can be seen in Table 4. Summarising

TABLE 2 | Fit measures for all models.

Model	No. of factors	χ^2	df	RMSEA	CFI	TLI	SRMR
Life satisfaction + anxiety + depression + worries	4	1372.45	38	0.05	0.97	0.96	0.03
Life satisfaction + PHQ-4 + worries	3	2259.56	41	0.06	0.95	0.94	0.03
Combined PHQ-4/worries + life satisfaction	2	6231.43	43	0.10	0.88	0.84	0.07
Single wellbeing factor	1	20825.19	44	0.19	0.58	0.47	0.16

PHQ-4: Patient Health Questionnaire—4 (combined anxiety and depression screening)

TABLE 3 | Fit measures for each wave and country—three factors model.

Wave	Country	χ2	df	RMSEA	CFI	TLI	SRMR
1	Austria	124.00	41	0.05	0.98	0.97	0.03
1	Brazil	286.47	41	0.07	0.96	0.95	0.03
1	Germany	294.59	41	0.08	0.94	0.92	0.05
1	Russia	110.23	41	0.05	0.98	0.97	0.04
1	UK	364.62	41	0.10	0.93	0.90	0.04
2	Austria	150.35	41	0.06	0.97	0.96	0.03
2	Brazil	256.32	41	0.06	0.97	0.96	0.03
2	Germany	287.23	41	0.08	0.95	0.93	0.04
2	Russia	138.42	41	0.06	0.97	0.96	0.04
2	UK	432.87	41	0.11	0.92	0.89	0.04
3	Austria	173.81	41	0.07	0.97	0.96	0.03
3	Brazil	331.30	41	0.07	0.96	0.95	0.03
3	Germany	218.38	41	0.07	0.96	0.95	0.04
3	Russia	117.65	41	0.05	0.98	0.97	0.03
3	UK	238.00	41	0.08	0.96	0.95	0.03

TABLE 4 | Invariance — three factors model.

	Group	ChiSq	df	CFI	RMSEA	GAMMA	MNCI	Δ CFI	Δ RMSEA	Δ GAMMA	Δ MNCI
Invariance across waves within each country											
Configural	Austria	448.15	123	0.98	0.06	0.97	0.95				
Metric	Austria	471.93	139	0.98	0.05	0.97	0.95	0.00	0.00	−0.00	−0.00
Scalar	Austria	573.87	155	0.97	0.06	0.97	0.94	0.01	−0.00	0.00	0.02
Configural	Brazil	874.09	123	0.96	0.07	0.96	0.93				
Metric	Brazil	897.35	139	0.96	0.06	0.96	0.93	0.00	0.00	−0.00	0.00
Scalar	Brazil	946.08	155	0.96	0.06	0.96	0.92	0.00	0.00	−0.00	0.00
Configural	Germany	800.20	123	0.95	0.08	0.94	0.91				
Metric	Germany	820.90	139	0.95	0.07	0.95	0.91	0.00	0.00	−0.01	0.00
Scalar	Germany	928.48	155	0.95	0.07	0.95	0.89	0.01	−0.00	0.00	0.01
Configural	Russia	366.30	123	0.98	0.05	0.97	0.96				
Metric	Russia	391.30	139	0.98	0.05	0.98	0.95	0.00	0.00	−0.00	0.00
Scalar	Russia	422.54	155	0.97	0.05	0.98	0.95	0.00	0.00	−0.00	0.00
Configural	UK	1035.49	123	0.94	0.10	0.92	0.86				
Metric	UK	1042.10	139	0.94	0.09	0.93	0.86	−0.00	0.01	−0.01	−0.00
Scalar	UK	1109.78	155	0.94	0.09	0.93	0.85	0.00	0.00	−0.00	0.01
Invariance across countries within each wave											
Configural	Wave 1	1179.90	205	0.96	0.07	0.95	0.92				
Metric	Wave 1	1300.79	237	0.95	0.07	0.95	0.91	0.00	0.00	−0.00	0.01
Scalar	Wave 1	2226.43	269	0.91	0.09	0.93	0.84	0.04	−0.02	0.03	0.07
Configural	Wave 2	1265.19	205	0.96	0.08	0.95	0.91				
Metric	Wave 2	1374.80	237	0.95	0.07	0.95	0.91	0.00	0.00	−0.00	0.00
Scalar	Wave 2	2397.96	269	0.91	0.09	0.92	0.82	0.04	−0.02	0.03	0.08
Configural	Wave 3	1079.14	205	0.97	0.07	0.96	0.93				
Metric	Wave 3	1225.56	237	0.96	0.07	0.96	0.92	0.00	0.00	−0.00	0.01
Scalar	Wave 3	2176.28	269	0.93	0.09	0.93	0.84	0.04	−0.02	0.03	0.08

the main patterns, we observed metric invariance across cultures at each wave and scalar invariance across time within each cultural sample. When separating the PHQ-4 scale into anxiety and depression, we observed identical results, available in Table S4.

3.3 | Wellbeing Differences Over Time

Using one-way anovas within each country and Bonferroni correction, we only found differences in Austria. People reported significantly more worries at time point 2 compared to time point 1 and 3: $F(2, 2235) = 9.037$, $p = 0.001$; W1 > W2; W2 < W3. Also, Life Satisfaction in Austria at time point 2 and 3 was significantly lower compared to time 1: $F(2, 2235) = 8.316$, $p = 0.003$; W1 > W2; W1 > W3. All other comparisons were not significantly different from each other (all $p > 0.05$). Figure 1 shows differences across waves.

3.4 | Wellbeing Correlations Across Cultures

Correlations between wellbeing variables are reported in Table 5 separately for each wave and country. Since there are three wellbeing correlations for each of the three time points (total of 45 correlations), and we can compute pairwise differences in these correlations between any two countries in our sample (90 comparisons), we decided to use Bonferroni adjustment to reduce the possibility of spurious effects. Out of the 90 possible pairwise differences, 8 comparisons were statistically significant after Bonferroni correction. The pattern suggested that the PHQ-4 and life satisfaction associations were weaker in the two lower income countries Brazil and Russia compared to the UK, Austria and Germany at time point 2 (Russia and UK, $Z = -5.080$, $p < 0.001$; Russia and Germany, $Z = -3.626$, $p = 0.025$, Russia and Austria, $Z = 5.036$, $p < 0.001$; Brazil and Austria, $Z = 4.696$, $p < 0.001$; Brazil and UK, $Z = -4.749$, $p < 0.001$). This relative disconnection of PHQ-4 and life satisfaction continued at time point 3 in Russia and Brazil compared to Austria (Brazil, $Z = 4.139$, $p = 0.003$; Russia, $Z = 3.666$, $p = 0.022$). The correlation between worries and life satisfaction was also weaker in Brazil compared to Austria at time point 3 ($Z = 4.178$, $p = 0.002$).

4 | Discussion

We reported invariance patterns of three wellbeing measures from different disciplinary perspectives across three time points

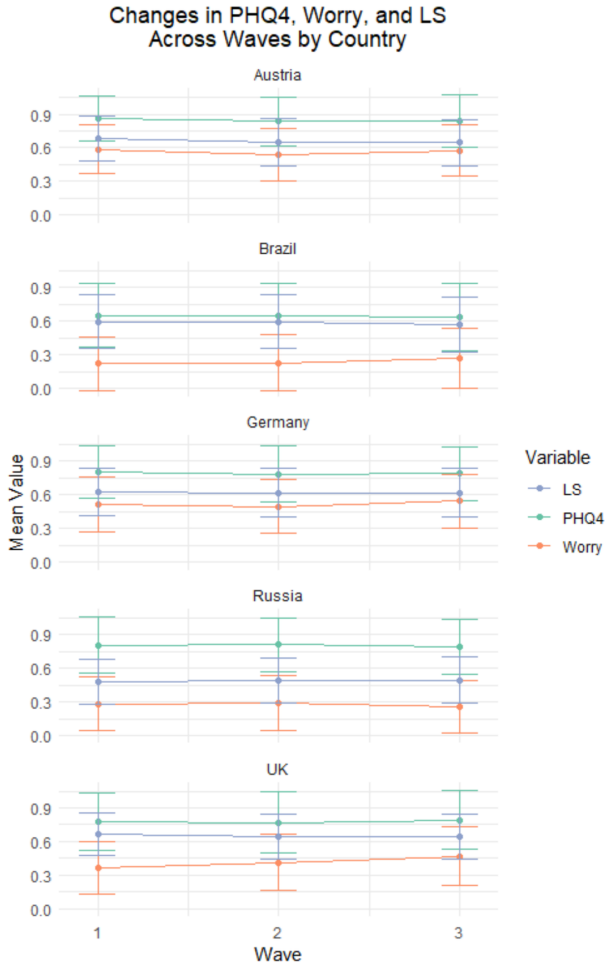

FIGURE 1 | Mean wellbeing scores across waves and countries.

and five countries during the recent COVID-19 pandemic. The invariance analysis suggests good cross-temporal replicability of the three wellbeing measures in five representative samples from distinct cultural regions, reaching scalar invariance longitudinally in each sample, and metric invariance across countries at each time point. The observation of good fit for a three factor structure suggests that pandemic worries, domain specific life satisfaction and the PHQ-4 as a general clinical screening tool are empirically distinct measures with moderate intercorrelations and that the measurement properties remained stable throughout the COVID-19 pandemic. Conceptually, these patterns imply that the three domains of wellbeing are conceptually distinct and capture related but conceptually different psychological processes that can be compared across time in each country. Exploring these effects over time, we found few differences in mean wellbeing over time. The relationship between measures also showed some differences between countries, that may be worth exploring further. We discuss these points in turn.

First, invariance tests across countries reached only metric invariance. An immediate implication is that means cannot be directly compared across samples nor can scores be interpreted as solely reflecting the specific wellbeing component. This means that researchers cannot interpret the average scores of the wellbeing measures as indicating the same underlying manifestation of wellbeing across different countries, because responses to items

TABLE 5 | Correlations between wellbeing variables.

Country	PHQ4_LS	PHQ4_Worry	LS_Worry
Wave 1			
Austria	0.50	0.31	0.25
Brazil	0.36	0.31	0.13
Germany	0.41	0.44	0.28
Russia	0.32	0.30	0.14
UK	0.43	0.40	0.20
Wave 2			
Austria	0.57	0.34	0.28
Brazil	0.35	0.28	0.13
Germany	0.49	0.36	0.25
Russia	0.30	0.26	0.20
UK	0.57	0.33	0.23
Wave 3			
Austria	0.59	0.38	0.32
Brazil	0.40	0.29	0.13
Germany	0.48	0.31	0.22
Russia	0.40	0.31	0.26
UK	0.50	0.31	0.21

by individuals in each sample are affected by other factors than the specific wellbeing construct. For example, even if two people from different groups have the same level of the underlying domain-specific life satisfaction, their scores may be shifted up or down due to some other variable that influences their responses in addition to their overall life satisfaction. Scalar non-invariance implies that there are other sources of variability that influence the intercepts of the items, and this exposes mean interpretations across samples to plausible alternative interpretations (e.g., response styles, different cognitive appraisals or thresholds). Hence, the observed scores do not necessarily reflect the individuals' true level of life satisfaction. These findings are in line with a number of recent studies suggesting that it is difficult to identify scalar invariance in culturally distinct samples (Emerson, Guhn, and Gadermann 2017; Giolo et al. 2022; Mónaco et al. 2022).

The lack of full invariance raises important questions for further theoretical elaboration and measurement development. We argue that addressing the invariance of psychological constructs across populations reveals key insights into whether invariance reflects measurement issues or actually cultural variations in how constructs are understood and manifested for further discussion see (Fischer and Rudnev 2024; Fischer et al. 2025). On the other hand, the evidence of scalar longitudinal invariance is encouraging, especially when considering some reports that supposedly stable individual difference measures changed their measurement properties during this period (Sutin et al. 2020). This guarantees that the relationships between the items and the underlying latent variables are the same across time points, and that the overall difficulty or ease of the items remains consistent. Achieving scalar invariance enables direct comparisons of mean scores and interpretations grounded in the assumed psychological construct. Therefore, it becomes possible to track changes

and especially recover in wellbeing during and after the pandemic. Nevertheless, since our data were collected after the onset of the pandemic, we cannot extrapolate to pre-pandemic wellbeing levels.

Investigating the stability of our wellbeing results through comparisons of nearly representative samples during the pandemic, the only major difference that survived after controlling for multiple comparisons was an increase in worries (note the reverse coding of the worry scale) during the second time point in Austria. This increase in worries may be explicable when considering that the data collection occurred soon after a major peak in relative mortality during the second wave in 2021. Also, life satisfaction was higher in the first data collection, and remained lower from the second to the third wave. Therefore, domain specific life satisfaction deteriorated in Austria during the COVID pandemic. Beyond these specific effects, the overall lack of change in mental health suggests that mental health remained relatively stable during the pandemic, at least when considering the individuals who participated in all three waves in each country. The evidence of scalar longitudinal invariance allows us to have greater confidence in the reported stability of wellbeing during the pandemic.

Finally, our analysis of the correlations between wellbeing variables suggested that life satisfaction in particular became more disconnected from clinical measures of depression and anxiety (and to some extent worries) in the lower income countries Brazil and Russia compared to the high-income countries, in particular Austria. These trends appeared more strongly during and after the second data collection wave. Therefore, clinical measures and domain specific life satisfaction as components of wellbeing became decoupled during the pandemic in lower-income societies in comparison to higher-income countries. Clinical measures may arouse different evaluations than life satisfaction, which may capture evaluations of the quality of life in specific domains. Such effects are worth exploring further, in particular as these measures are typically studied in isolation and our patterns may imply that positive and negative wellbeing effects may be more distinct in lower income countries. Measuring life satisfaction may provide little information on clinical states in majority world countries. More broadly, these findings also call into question the relative conceptual unity of subjective wellbeing across cultures and imply that wellbeing domain effects warrant greater theoretical attention.

4.1 | Future Research Directions and Limitations

Our study of wellbeing measurement invariance benefits from the size and characteristics of our sample, but we also acknowledge some constraints. A primary limitation stems from the fact that our data collection time points are not perfectly aligned across countries and may represent different stages of the COVID-19 pandemic in each country. A second limitation is the use of short scales, even though this is probably unavoidable within the constraints of population-level longitudinal studies due to cost–benefit considerations. Future research needs to explore options to incorporate more comprehensive instruments that capture wellbeing more holistically, including both domain specific and general measures of wellbeing. Finally, while our samples exhibit broader representation compared to the predominantly student-centric samples and European or English-speaking samples in psychological science, they may not be fully representative or include all world cultures. We were only able to include data from individuals who provided information at all three time points within the Values in Crisis project, which not only led to dropouts that may have introduced possible demographic biases compared to the overall population but also to the selection of specific populations across the world. Dealing with three time points, we believe imputation of missing data based on two observations is hazardous. Our results therefore have to be interpreted with this possible sampling bias in mind. Our sampling is nevertheless a major advantage compared to the typical student and small convenience samples that are dominant in psychology. Future studies could benefit from comparing even more diverse countries.

5 | Conclusion

We report evidence of longitudinal full score invariance of three wellbeing instruments, which is reassuring given recent evidence that wellbeing instruments may have been unstable or unreliable during major events such as the COVID-19 pandemic. We observed relative stability in wellbeing means over time, which suggests that wellbeing remained largely stable during the pandemic. In line with other research, we found only metric invariance at each wave across samples. This means that mean differences across samples are open to alternative interpretations and direct cross-cultural comparisons of wellbeing during the pandemic are not possible. Our data also suggest that clinical measures may have become dissociated from life satisfaction in low-income countries during the second wave of the pandemic.

Ethics Statement

All procedures performed in studies involving human participants were in accordance with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The research has received approval from a local ethics board (CAAE 30823020.6.0000.5249).

Consent

Informed consent was obtained from all individual adult participants included in the study.

Conflicts of Interest

The authors declare no conflicts of interest.

References

Aschauer, W., A. Seymer, O. Bodi-Fernandez, and M. Herzog. 2021. "Values in Crisis International (SUF edition)." *Database*. https://doi.org/10.11587/LIHK1L.

Beuke, C. J., R. Fischer, and J. McDowall. 2003. "Anxiety and depression: Why and How to Measure Their Separate Effects." *Clinical Psychology Review* 23, no. 6: 831–848. https://doi.org/10.1016/S0272-7358(03)00074-6.

Caro-Fuentes, S., and J. P. Sanabria-Mazo. 2024. "A Systematic Review of the Psychometric Properties of the Patient Health Questionnaire-4 in Clinical and Nonclinical Populations." *Journal of the Academy of*

Consultation-Liaison Psychiatry 65, no. 2: 178–194. https://doi.org/10.1016/j.jaclp.2023.11.685.

Caycho-Rodríguez, T., L. W. Vilca, P. D. Valencia, et al. 2023. "Is the Meaning of Subjective Well-Being Similar in Latin American Countries? A Cross-Cultural Measurement Invariance Study of the WHO-5 Well-Being Index During the COVID-19 Pandemic." *BioMed Central Psychology* 11, no. 1: 1–14. https://doi.org/10.1186/s40359-023-01149-8.

Cohen, P., J. Cohen, L. S. Aiken, and S. G. West. 1999. "The Problem of Units and the Circumstance for POMP." *Multivariate Behavioral Research* 34, no. 3: 315–346. https://doi.org/10.1207/S15327906MBR3403_2.

Diener, E. 1984. "Subjective Well-Being." *Psychological Bulletin* 95, no. 3: 542–575. https://doi.org/10.1037/0033-2909.95.3.542.

Diener, E., R. Emmons, R. Larsen, and S. Griffin. 1985. "The Satisfaction With Life Scale." *Journal of Personality Assessment* 49, no. 1: 71–75. https://doi.org/10.1207/s15327752jpa4901_13.

Emerson, S. D., M. Guhn, and A. M. Gadermann. 2017. "Measurement Invariance of the Satisfaction With Life Scale: Reviewing Three Decades of Research." *Quality of Life Research* 26, no. 9: 2251–2264. https://doi.org/10.1007/s11136-017-1552-2.

Fischer, R., and J. A. Karl. 2019. "A Primer to (Cross-Cultural) Multi-Group Invariance Testing Possibilities in R." *Frontiers in Psychology* 10, no. JULY: 1–18. https://doi.org/10.3389/fpsyg.2019.01507.

Fischer, R., J. A. Karl, M. Luczak-Roesch, and L. Hartle. 2025. "Why We Need to Rethink Measurement Invariance: The Role of Measurement Invariance for Cross-Cultural Research." *Cross-Cultural Research*: 1–33. https://doi.org/10.1177/10693971241312459.

Fischer, R., and M. Rudnev. 2024. "From MIsgivings to MIse-en-scène: The Role of Invariance in Personality Science." *European Journal of Personality*: 1–12. https://doi.org/10.1177/08902070241283081.

Fischer, R., and E. Van de Vliert. 2011. "Does Climate Undermine Subjective Well-Being? A 58-Nation Study." *Personality and Social Psychology Bulletin* 37, no. 8: 1031–1041. https://doi.org/10.1177/0146167211407075.

Giolo, S. R., R. C. F. Giordani, M. Zanoni da Silva, P. C. A. Dias, A. Estavela, and J. Ismael Mabuie. 2022. "Cross-Cultural Measurement Invariance of the Fear of COVID-19 Scale in Three Portuguese-Speaking Countries." *Journal of Health Psychology* 27, no. 13: 2997–3012. https://doi.org/10.1177/13591053221076578.

Hu, L., and P. M. Bentler. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives." *Structural Equation Modeling: A Multidisciplinary Journal* 6, no. 1: 1–55. https://doi.org/10.1080/10705519909540118.

Inglehart, R., and C. Welzel. 2005. *Modernization, Cultural Change, and Democracy: The Human Development Sequence*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790881.

Kerstis, B., F. Giannotta, P. v. H. Wågert, et al. 2021. "Changes in Mental Health and Views on Communication and Activities of Public Institutions Among Swedes During the COVID-19 Pandemic—A Cross-Sectional Repeated Measures Design." *Health* 9, no. 11: 1498. https://doi.org/10.3390/healthcare9111498.

Lance, C. E., A. G. Mallard, and A. C. Michalos. 1995. "Tests of the Causal Directions of Global-Life Facet Satisfaction Relationships." *Social Indicators Research* 54, no. 1: 69–92. https://doi.org/10.1007/BF01078968.

Lanz, M., R. Caliciuri, R. Iafrate, C. Regalia, R. Rosnati, and A. Sorgente. 2023. "Covid-Related Stress in the Financial, Relational and Health Domains. Which Longitudinal Effects on Present and Future Perception?" *Psychological Reports*: 1–27. https://doi.org/10.1177/00332941231161790.

Leitgöb, H., D. Seddig, T. Asparouhov, et al. 2023. "Measurement invariance in the social sciences: Historical Development, Methodological Challenges, State of the Art, and Future Perspectives." *Social Science Research* 110: 102805. https://doi.org/10.1016/j.ssresearch.2022.102805.

Leonardi, F., L. Spazzafumo, and F. Marcellini. 2005. "Subjective Well-Being: The Constructionist Point of View. A Longitudinal Study to Verify the Predictive Power of Top-Down Effects and Bottom-Up Processes." *Social Indicators Research* 70, no. 1: 53–77.

Locke, J. 2004. *An Essay Concerning Humane Understanding, Volume II*. Project Gutenberg. http://www.gutenberg.org/ebooks/10615.

Löwe, B., I. Wahl, M. Rose, et al. 2010. "A 4-Item Measure of Depression and Anxiety: Validation and Standardization of the Patient Health Questionnaire-4 (PHQ-4) in the General Population." *Journal of Affective Disorders* 122, no. 1–2: 86–95. https://doi.org/10.1016/j.jad.2009.06.019.

Maassen, E., E. D. D'Urso, M. A. L. M. van Assen, M. B. Nuijten, K. De Roover, and J. M. Wicherts. 2023. "The Dire Disregard of Measurement Invariance Testing in Psychological Science." *Psychological Methods*: 1–14. https://doi.org/10.1037/met0000624.

Manão, A., A. T. Pereira, C. Cabacos, A. P. Amaral, M. J. Soares, and A. Macedo. 2022. "Worries During the COVID-19 Pandemic—Which Were the Most Prevalent and Disturbing?" *European Psychiatry* 65, no. S1: S191–S192. https://doi.org/10.1192/j.eurpsy.2022.504.

Mónaco, E., K. Schoeps, S. Valero-Moreno, et al. 2022. "Cross-Cultural Validation of the Worries about COVID-19 and its Consequences Scale (W-COV) in Adolescents and Young People." *Archives of Psychiatric Nursing* 40: 158–166. https://doi.org/10.1016/j.apnu.2022.07.016.

Nakazato, N., U. Schimmack, and S. Oishi. 2011. "Effect of Changes in Living Conditions on Well-Being: A Prospective Top–Down Bottom–Up Model." *Social Indicators Research* 100, no. 1: 115–135. https://doi.org/10.1007/s11205-010-9607-6.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Rutkowski, L., and D. Svetina. 2014. "Assessing the Hypothesis of Measurement Invariance in the Context of Large-Scale International Surveys." *Educational and Psychological Measurement* 74, no. 1: 31–57. https://doi.org/10.1177/0013164413498257.

Sawicki, A. J., M. Żemojtel-Piotrowska, J. M. Balcerowska, et al. 2022. "The Fear of COVID-19 Scale: Its Structure and Measurement Invariance Across 48 Countries." *Psychological Assessment* 34, no. 3: 294–310. https://doi.org/10.1037/pas0001102.

Schwartz, S. H., L. Sagiv, and K. Boehnke. 2000. "Worries and Values." *Journal of Personality* 68, no. 2: 309–346. https://doi.org/10.1111/1467-6494.00099.

Steinmayr, R., P. Paschke, and L. Wirthwein. 2022. "Elementary School Students' Subjective Well-Being Before and During the COVID-19 Pandemic: A Longitudinal Study." *Journal of Happiness Studies* 23, no. 6: 2985–3005. https://doi.org/10.1007/s10902-022-00537-y.

Sutin, A. R., M. Luchetti, D. Aschwanden, et al. 2020. "Change in Five-Factor Model Personality Traits During the Acute Phase of the Coronavirus Pandemic." *PLoS One* 15, no. 8: e0237056. https://doi.org/10.1371/journal.pone.0237056.

Welzel, C., K. Boehnke, J. Delhey, et al. 2023. "Values in Crisis—A Crisis of Values? Moral Values and Social Orientations Under the Imprint of the Corona pandemic. GESIS, Cologne. ZA7989 Data file Version 1.0.0." https://doi.org/10.4232/1.14148.

Wicke, F. S., L. Krakau, B. Löwe, M. E. Beutel, and E. Brähler. 2022. "Update of the Standardization of the Patient Health Questionnaire-4 (PHQ-4) in the General Population." *Journal of Affective Disorders* 312: 310–314. https://doi.org/10.1016/j.jad.2022.06.054.

Yap, S., A. Lee, L. J. Ji, Y. Li, and Y. Dong. 2021. "Cultural Differences in People's Psychological Response to COVID-19." *Frontiers in Psychology* 12: 1–13. https://doi.org/10.3389/fpsyg.2021.636062.

Zacher, H., and C. W. Rudolph. 2021. "Individual Differences and Changes in Subjective Wellbeing During the Early Stages of the COVID-19 Pandemic." *American Psychologist* 76, no. 1: 50–62. https://doi.org/10.1037/amp0000702.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.