




Validation and Measurement Invariance of the Personal Financial Wellness Scale

A Multinational Study in 7 Countries

Eike K. Buabang^{1,2} , Sarah Ashcroft-Jones³, Celia Esteban Serna⁴, Katarina Kastelic^{5,6}, Jakob Kveder⁵, Amanda Lambertus⁷, Tasja S. Müller⁸, and Kai Ruggeri⁹

¹Research Group of Quantitative Psychology and Individual Differences, KU Leuven, Belgium

²Center for Social and Cultural Psychology, KU Leuven, Belgium

³Department of Experimental Psychology, University of Oxford, UK

⁴Division of Psychology & Language Sciences, University College London, UK

⁵Department of Psychology, Faculty of Social Sciences, Lund University, Sweden

⁶Novo Mesto Health Center, Novo Mesto, Slovenia

⁷Institute of Psychiatry Psychology & Neuroscience, King's College London, UK

⁸Faculty of Behavioural and Social Sciences, University of Groningen, The Netherlands

⁹Department of Health Policy and Management, Mailman School of Public Health, Columbia University, New York City, NY, USA

Abstract: In 2020, 17.1% of the population in the European Union was at risk of poverty (Eurostat, 2021). Poverty is often assessed using objective measures such as absolute and relative income levels. However, different individuals may experience different levels of financial stress at the same income level. Therefore, it is crucial to have measures that capture the subjective components of poverty. In this multinational study, we tested the validity and measurement invariance of the Personal Financial Wellness (PFW) scale across six European countries (Germany, Italy, the Netherlands, Slovenia, Spain, and the UK) and the US, and six languages (German, Italian, Dutch, Slovenian, Spanish, and English). Results provided mixed evidence for the fit of the expected one-factor structure. Exploration of a modified one-factor structure indicated an improved fit. The scale showed excellent reliability, and convergent and discriminant validity. This suggests that the PFW scale captures subjective financial stress and is a dependable self-report measure. Measurement invariance testing of the modified one-factor model showed metric invariance across Slovenia, Spain, the UK, and the US. Given that scalar invariance was not achieved and the invariance testing was based on an exploratory model, we do not advise the use of the scale for comparisons between countries.

Keywords: Personal Financial Wellness Scale, validation, measurement invariance, financial stress, poverty



In 2020, over 75 million people in the European Union (EU) were at risk of poverty, meaning they had 60% or less of the median disposable income (adjusted for household size) in their country (Eurostat, 2021). This at-risk-of-poverty threshold defines poverty based purely on the objective measure of income. However, this definition fails to address how individuals perceive their financial situation. For example, the same income can be perceived differently

by different individuals (Prawitz et al., 2006), creating unique subjective experiences of financial stress. Sommet and colleagues (2018) found that the perception of income is only moderately correlated with actual income level and that it is the perception itself that relates to increased rates of unhappiness and poor mental health. Therefore, to understand financial stress, it is critical to have measures that accurately capture these subjective experiences.

Subjective financial stress has been measured in a number of ways. One approach involves defining a subjective poverty line in economic space (Ravallion, 2012). This can be achieved by asking people to assess their financial situation in relation to others by placing themselves on a

rung of an economic ladder, where the bottom two rungs are indicative of subjective poverty (e.g., Ravallion & Lokshin, 2002). Alternatively, researchers can ask what people consider to be the minimum income they require to “make ends meet” and the subjective poverty line is drawn at their minimum required income (e.g., Hagenaars & van Praag, 1985). Another approach simply asks people if they are able to satisfy their basic needs, thereby measuring financial strain (e.g., Kahn & Pearlin, 2006). Finally, subjective experiences can be captured by measuring affective states, which include feelings of worry, stress, and anxiety about one’s financial situation (e.g., Shapiro & Burchell, 2012). Individually, these measures give insight into particular aspects of subjective financial stress, however, to gain a more complete picture a more holistic measure is required. A candidate scale to measure subjective financial stress in a more complete manner is the Personal Financial Wellness Scale[®] (PFW Scale; Prawitz et al., 2006).

The PFW scale (Table 1) consists of eight items measuring financial stress and well-being (Prawitz et al., 2006). It was initially developed under the name InCharge Financial Distress/Financial Well-being (IFDFW) scale and was validated in the United States in the general population and in consumer credit counseling clients (Prawitz et al., 2006). The scale includes items that measure different aspects of financial stress, such as financial strain and affective states. Lower scores on the scale indicate lower levels of financial well-being and higher levels of financial stress. Different PFW scores have been associated with different financial behaviors. For example, improved budgeting and saving when financial well-being is high, or risky credit card usage and compulsive buying when financial stress increases (Gutter & Copur, 2011). Furthermore, people who contacted a credit counseling agency scored lower on the PFW scale (Prawitz et al., 2006). Given the breadth of the PFW scale and its relation to real-life financial behaviors, in this study, we aimed to test the validity and measurement invariance of the PFW scale across six European countries (Germany, Italy, the Netherlands, Slovenia, Spain, the United Kingdom), and the United States.

The validity and reliability of the scale were first established in the US. Prawitz and colleagues (2006) provide a detailed account of how the scale was initially developed. A principal component analysis (PCA) of the final 8-item scale indicated that the items measured one factor, explaining 78.9% of the variance. The loadings for each item on the factor were between .83 to .93. The scale showed high internal consistency, indicated by a Cronbach’s α of .96. The validity and reliability of the scale were replicated in a college sample from the US (Gutter & Copur, 2011). Using a PCA, 64% of the total variance was explained by one factor, with factor loadings ranging from .69 to .90 and a Cronbach’s α coefficient of .91. A further study by Nielsen (2010) using

confirmatory factor analysis (CFA) indicated that the one-factor solution explained just 56.6% of the variance. Cronbach’s α coefficient was between .89 and .90 depending on the handling of missing data. Nielsen (2010) proposed an alternative two-factor structure for the scale distinguishing between items that measure objective (Items 1, 2, 3, 8) and subjective (Items 4–7) financial stress. A forced CFA for the two-factor model demonstrated a reasonable fit, however, the eigenvalue of the objective factor did not exceed the cut-off value of 1. There may also be theoretical reasons against adopting a two-factor model as items from both subscales are still a subjective assessment of financial stress. Collectively these findings establish the validity and reliability of the PFW scale in US and indicate that a one-factor structure is a reasonable fit across multiple samples.

Outside the US context, Kamaluddin and colleagues (2018) translated the scale to Malay and examined validity and reliability in Malaysia. They examined the scale using an exploratory factor analysis (EFA) with PCA as the extraction method. They found a one-factor structure, explaining 65.28% of the total variance. The factor loadings were between .62 and .88. Cronbach’s α coefficient was .92. In the current study, we aimed to investigate the validity of the PFW scale in the European context. Our sample includes six countries from different areas of Europe (Germany, Italy, the Netherlands, Slovenia, Spain, and the UK). This study involved the translation of the scale into five languages (German, Italian, Dutch, Slovenian, and Spanish). Beyond the European context, we also attempted to replicate the one-factor structure in the US. We planned to conduct a multi-group CFA, testing the structure and measurement invariance of the scale across these seven countries.

Arguably, the role of financial stress differs between the countries included in the current study. For example, compared to Europe, income inequality is higher in the US (Blanchet et al., 2020). Christelis and colleagues (2017) find that more US households have debt compared to European households and that the debt is relatively higher. According to the Stress in America report (American Psychological Association, 2021), 61% of Americans report that money is a significant source of stress. Based on these differences, we might expect higher baseline financial stress levels in the US compared to the European countries. There are other differences between the countries, for instance, the extent to which they are individualistic or collectivistic cultures (Hofstede et al., 2010) which may influence subjective financial stress. While we acknowledge these differences, we did not make specific predictions for the differences between the countries.

To account for these differences, and to allow for differences in how the one-factor model of the PFW scale fits across countries, we planned to conduct separate CFAs for each country. We planned to then conduct a multi-group

Table 1. Personal Financial Wellness Scale[®]

Item	Score (1–10)
What do you feel is the level of your financial stress today?	1 = Overwhelming stress 4 = High stress 7 = Low stress 10 = No stress at all
How satisfied are you with your present financial situation?	1 = Completely dissatisfied 4 = Somewhat dissatisfied 7 = Somewhat satisfied 10 = Completely satisfied
How do you feel about your current financial condition?	1 = Feel overwhelmed 4 = Sometimes feel worried 7 = Not worried 10 = Feel comfortable
How often do you worry about being able to meet normal monthly living expenses?	1 = All the time 4 = Sometimes 7 = Rarely 10 = Never
How confident are you that you could find the money to pay for a financial emergency that costs about \$1,000?	1 = No confidence 4 = Little confidence 7 = Some confidence 10 = High confidence
How often does this happen to you? You want to go out to eat, go to a movie or do something else and don't go because you can't afford to?	1 = All the time 4 = Sometimes 7 = Rarely 10 = Never
How frequently do you find yourself just getting by financially and living paycheck to paycheck?	1 = All the time 4 = Sometimes 7 = Rarely 10 = Never
How stressed do you feel about your personal finances in general?	1 = Overwhelming stress 4 = High stress 7 = Low stress 10 = No stress at all

Note. Each item is scored from 1 to 10 and the total is divided by 8. Lower scores indicate lower financial well-being. Scale used with permission.

CFA and measurement invariance testing, excluding those countries with an unacceptable model fit. Further, we examined convergent validity by assessing the extent to which the PFW scale is related to the Perceived Stress Scale (PSS; Cohen et al., 1983). Convergent correlations between the PSS and objective financial condition have previously been found in various countries (e.g., Croatia (Fazlić et al., 2012) and Spain (Vallejo et al., 2018)). Lower scores on the PFW scale indicate more financial stress. Therefore, we expected a strong negative correlation ($> -.50$) between the PFW scale and PSS scores across all countries.

Finally, we assessed discriminant validity by relating the PFW scale to the Ten Item Personality Inventory (TIPI; Gosling et al., 2003). Financial stress has been found to negatively correlate with conscientiousness and positively with neuroticism (Xu et al., 2015). The relations between

financial stress and extraversion, openness to experience, and agreeableness are less consistent. In line with these findings, we expected moderately negative correlations ($-.30$ to $-.50$) between conscientiousness and PFW scores, and moderately positive correlations ($.30$ to $.50$) between neuroticism and PFW scores. We expected low correlations for the remaining three traits. For both convergent and discriminant validity, correlations were also compared between the US and each European country.

Methods

Participants

We aimed to recruit comparable participant samples of at least 200 participants per country to meet the

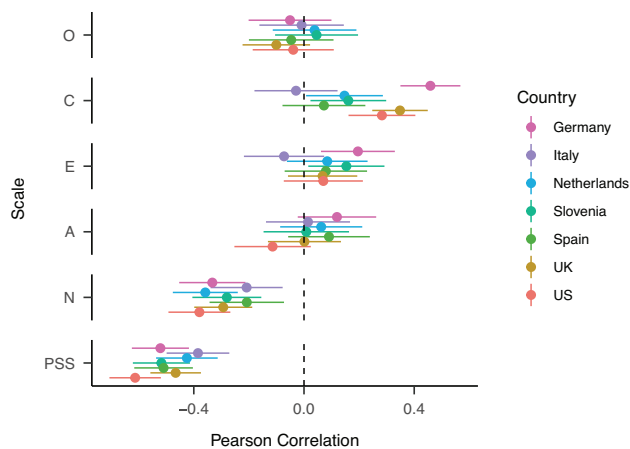


Figure 1. Correlations between the PFW and the TIPI traits and PSS scores country. Correlations between the PFW and the TIPI and PSS scales are shown. Correlations were computed within each country. The five personality traits assessed by the TIPI subscales are denoted by the letters OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism).

recommended minimum sample size for multi-group CFA (Fischer & Karl, 2019). We recruited a total of 1,877 participants. Those who did not consent to or complete the study were excluded from the sample prior to data cleaning leaving 1,553 participants. Participants were excluded based on the time they took to complete the study. Participants who took less than 25% of the mean duration per sampling type were excluded ($n = 3$) leaving a final sample of $N = 1,550$ participants that were included in the analyses across the seven countries (Germany: ($n = 205$), Italy: ($n = 213$), Netherlands: ($n = 207$), Slovenia: ($n = 212$), Spain: ($n = 203$), UK: ($n = 288$), US: ($n = 222$). Full demographic information is available in the Supplementary Materials, Table 1 and Figure 1: <https://osf.io/gevmr/>).

Data missingness was also checked, overall $< 0.1\%$ was missing and 100% of the PFW data was present. Some data was missing for TIPI items, these participants ($n = 22$) were included in the main analysis but underwent listwise deletion for the validity analyses.

Procedure

The survey was created using Qualtrics software. Participants were recruited in two ways. First, using the Prolific platform on which participants were compensated according to the minimum fee. Second, using snowball sampling on social media. The proportions of participants recruited

using each method varied across countries. For example, all UK and US participants were sampled via snowball sampling, while all participants from the Netherlands were sampled using Prolific. For the other countries, the proportions sampled on Prolific or via snowball sampling were mixed (Prolific Proportion: Germany = 0.78, Italy = 0.80, Slovenia = 0.43, Spain = 0.88). At the start of the survey, participants were informed of the confidentiality and anonymity of the collected data, and the voluntary nature of participation. Next, they were presented with a series of measures including the PFW scale, PSS, and TIPI. The validation of the PFW scale is part of a larger study (the full list of measures is available in the Supplementary Materials, <https://osf.io/gevmr/>). Finally, participants were debriefed and provided with the contact details of the research team.

Materials

The Personal Financial Wellness Scale (PFW scale; Prawitz et al., 2006)

Each item is scored from 1 to 10 and the total score is divided by 8, the total number of items. Lower scores indicate higher financial distress and lower financial well-being. (See Table 1 for item details, scale used with permission¹).

Perceived Stress Scale (PSS; Cohen, 1988)

The 10-item PSS (Cohen, 1988) was used because the psychometric properties were found to be superior to those of the 14-item PSS (Lee, 2012). This version includes six negatively worded items (e.g., “In the last month, how often have you felt nervous and “stressed”?”) and four positively worded items (e.g., “In the last month, how often have you felt that things were going your way?”). Each item uses a 5-point Likert format ranging from 0 = *never* to 4 = *very often*. Total scores are obtained by reverse coding the four positively worded items and summing across all scale items. Scores will range from 0 to 40 points, where higher scores indicate higher perceived stress.

Ten Item Personality Inventory (TIPI; Gosling et al., 2003)

The TIPI is a short ten-item measure of the Big-Five personality dimensions (extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience). Each dimension is measured with two items on a 7-point scale. The scale asks to report agreement (*disagree strongly* to *agree strongly*) to items such as “I see myself as extraverted, enthusiastic”.

¹ © Copyright by Personal Finance Employee Education Fund (<http://www.pfeef.org>), E. Thomas Garman, and/or John Hoffmire. (1) E. Thomas Garman, Fellow and Professor Emeritus, Virginia Tech University; 1761 Pennecamp Drive, The Villages, FL 32162, USA; Tel.: 352-205-4283; E-mail: ethomasgarman@yahoo.com. (2) John Hoffmire, Emeritus, University of Wisconsin-Madison, 12 The Paddock, Oxford, OX1 5SB, UK; Tel.: +44 1865 701914; Mobile: +44 78844 72169; E-mail: hoffmire@wisc.edu. Obtain permission for use from any of the above. All rights reserved.

Translation Procedure

The translation was done in three steps. First, native speakers translated the scales from their original English version into the target languages (German, Italian, Dutch, Slovenian, and Spanish). Second, in each country, a set of bilingual speakers who were unaware of the scale and the purpose of the study conducted a back-translation. Finally, for each language, a bilingual psychometric expert evaluated the translation of the scale, which was compared to the original version.

Data Analysis

We used R (Version 4.2.0; R Core Team, 2022) and the R-packages *cocor* (Version 1.1.3; Diedenhofen & Musch, 2015), *dplyr* (Version 1.0.8; Wickham et al., 2022), *Hmisc* (Version 4.6.0; Harrell Jr., 2022), *kableExtra* (Version 1.3.4; Zhu, 2021), *knitr* (Version 1.39; Xie, 2015), *lavaan* (Version 0.6.11; Rosseel, 2012), *lordif* (Version 0.3.3; Choi et al., 2016), *Olivoto2020* (Olivoto & L'ucio, 2020), *papaja* (Version 0.1.0.9999; Aust & Barth, 2020), *semTools* (Version 0.5.5; Jorgensen et al., 2021), *tidyverse* (Version 1.3.1; Wickham et al., 2019), *userfriendlyscience* (Version 0.7.2; Peters, 2017), and *visdat* (Version 0.5.3; Tierney, 2017) for all our analyses.

Testing measurement invariance of the PFW scale follows the guidelines by Fischer and Karl (2019) and used the *lavaan* package in R. First, we conduct separate CFAs for each country to examine the one-factor model fit in each country. The cut-off scores for the interpretation of each index are presented in Table 2 (adapted from Greiff & Allen, 2018).

To assess internal consistency, we report Cronbach's α and ω (Crutzen & Peters, 2017). Alpha and omega are obtained using the *userfriendly science* package in R and the code provided by Peters (2014). Confidence intervals for α and ω are computed through bootstrapping (1,000 samples). Observation of poor fit in any country results in exclusion from the multi-group CFA and measurement invariance testing. Excluded countries are subject to further examination to explore reasons for their poor fit.

We plan to run a multi-group CFA testing the one-factor model across the remaining countries. The procedure for the multi-group CFA is as follows: after assessing the model fit indices, we test configural, metric, scalar, and strict invariance. We accept a more constrained model if the Δ CFI between a less constrained and more constrained model is ≤ 0.01 . If the Δ CFI is > 0.01 , we test the partial invariance of the last model that was not invariant. We check which parameter would result in the largest Δ CFI if we allow it to vary across countries. The possible variation of this parameter is also evaluated on a theoretical level.

Table 2. Model fit cut-off scores

Item	CFI	TLI	RMSEA	SRMR	α and ω
Excellent	> 0.95	> 0.95	$< .06$	$< .06$	> 0.90
Good	> 0.92	> 0.92	$< .07$	$< .07$	> 0.85
Acceptable	> 0.89	> 0.89	$< .08$	$< .08$	> 0.80
Questionable	> 0.86	> 0.86	$< .09$	$< .09$	> 0.75
Poor	> 0.83	> 0.83	$< .10$	$< .10$	> 0.70
Unacceptable	< 0.83	< 0.83	$> .10$	$> .10$	< 0.70

Note. Model fit cut-off values adapted from Greiff and Allen (2018).

A model in which this parameter is free from the constraint is compared against the last model that was invariant. Δ CFI is assessed again. If Δ CFI is ≤ 0.01 , we accept a partial invariant model. If it is > 0.01 , freeing up another parameter will be considered.

Results

Registered Analyses

The one-factor model was first run in each country separately (see Table 3). The CFI was excellent ($> .95$) in Germany and the US, it was good ($> .92$) in Italy, the Netherlands, and Slovenia, and acceptable ($> .89$) in Spain and the UK. The TLI was good ($> .92$) in Germany, Slovenia, and the US, it was acceptable ($> .89$) in Italy and the Netherlands, and questionable ($> .86$) in Spain and the UK. The RMSEA was poor for each country with values ranging above .10. The SRMR was excellent ($< .06$) in all countries. Similarly, the inter-item reliability measures were excellent ($> .90$) in all countries. Based on the unacceptable RMSEA values, no countries met the cut-off for inclusion in a multi-group CFA and subsequent measurement invariance testing using the registered one-factor model.

To examine the poor fit of the registered one-factor model, individual item loadings were then assessed within each country. The item loadings varied across and within countries with values ranging between 0.67–0.86 in Germany, 0.66–0.82 in Italy, 0.64–0.84 in the Netherlands, 0.71–0.86 in Slovenia, 0.68–0.91 in Spain, 0.69–0.89 in the UK, and 0.74–0.91 in the US. Across all countries, apart from the UK (where it was the second lowest), Item 5 (“Confidence regarding financial emergency”) had the lowest loading. Across four countries (Italy, Spain, the UK, and the US) the four lowest loading items were Items 4–7. In the Netherlands, the three lowest-loading items were Items 5–7, and in Slovenia, the two lowest-loading items were Items 5–6. These items (Items 4–7 inclusive) appear to load poorly to the one-factor model across countries. While this does not mean that these items necessarily load to a second factor, they have previously been suggested to load to a

Table 3. Model fit indices for individual country CFAs based on one-factor model of the PFWs

Country	N	χ^2	df	CFI	TLI	RMSEA	SRMR	α	95% CI		ω	95% CI	
									LL	UL		LL	UL
Germany	205	75	20	0.95	0.93	0.12	0.04	0.92	0.91	0.94	0.93	0.91	0.94
Italy	213	95	20	0.93	0.90	0.13	0.05	0.91	0.89	0.93	0.91	0.89	0.93
Netherlands	207	90	20	0.93	0.90	0.13	0.05	0.91	0.89	0.92	0.91	0.89	0.93
Slovenia	212	84	20	0.95	0.93	0.12	0.04	0.93	0.91	0.94	0.93	0.91	0.94
Spain	203	144	20	0.90	0.87	0.18	0.06	0.93	0.91	0.94	0.93	0.91	0.94
UK	288	184	20	0.91	0.88	0.17	0.05	0.94	0.92	0.95	0.93	0.92	0.95
US	222	100	20	0.95	0.93	0.13	0.04	0.95	0.94	0.96	0.95	0.94	0.96

Note. Alpha and omega reported with bootstrapped 95% confidence intervals (CI). LL indicates the lower limit and UL the upper limit of these intervals. Results are reported rounded to 2 decimal places, however, unrounded values are used to assess model fit indices.

Table 4. Factor loading per item per country

Item	Germany	Italy	The Netherlands	Slovenia	Spain	UK	US
Level of financial stress today	0.69	0.82	0.75	0.80	0.91	0.84	0.89
Satisfaction with present financial situation	0.82	0.82	0.78	0.77	0.81	0.84	0.84
Feelings about current financial situation	0.86	0.80	0.84	0.85	0.88	0.89	0.91
Worry about monthly living expenses	0.81	0.72	0.79	0.82	0.76	0.69	0.80
Confidence regarding financial emergency	0.67	0.66	0.64	0.71	0.68	0.75	0.74
Can't afford to go out	0.82	0.69	0.72	0.73	0.69	0.80	0.81
Paycheck to paycheck	0.82	0.76	0.71	0.82	0.77	0.81	0.82
Stress about finances in general	0.83	0.82	0.82	0.86	0.88	0.88	0.91

Note. Standardized factor loading per item of the PFW scale per country.

subscale of the PFW: “Objective Financial Wellbeing” as part of a two-factor model (Nielsen, 2010). However, the two-factor model did not meet the eigenvalue cut-off required to establish this subscale (Nielsen, 2010). This pattern of results is discussed in more detail in the exploratory analyses section below. All standardized individual item loading per country countries are reported in Table 4.

Convergent Validity

To assess convergent validity, correlations between scores on the PFW scale and PSS were calculated (see Figure 1). Lower scores on the PFW scale indicate more financial stress. Therefore, we expected a strong negative correlation (> -0.50) between the PFW scale and PSS scores. In line with our prediction, results indicated significantly high negative correlations in the US ($r = -0.61, p < .001$), in Spain ($r = -0.51, p < .001$), in Germany ($r = -0.52, p < .001$), and in Slovenia ($r = -0.52, p < .001$). Results further indicated significant moderate to high negative correlations in the UK ($r = -0.47, p < .001$), in the Netherlands ($r = -0.43, p < .001$), and in Italy ($r = -0.38, p < .001$).

Compared to the US, the correlation was significantly lower in the UK, ($z = -2.31, p = .021$), in the Netherlands, ($z = -2.65, p = .008$), and in Italy, ($z = -3.16, p = .002$).

The correlations in the other countries were not significantly different from the US (all z s < -1.44).

Discriminant Validity

Discriminant validity was assessed by examining correlations with the five traits measured with the TIPI (see Figure 1). We expected moderately negative correlations ($-.30$ to $-.50$) between conscientiousness and PFW scores, and moderately positive correlations ($.30$ to $.50$) between neuroticism and PFW scores. We expected low correlations for the remaining three traits.

Conscientiousness

In line with our prediction, results indicated significant moderate to high positive correlations in Germany ($r = 0.46, p < .001$), and in the UK ($r = 0.35, p < .001$). Results further indicated significant small to moderate correlations in the Netherlands ($r = 0.15, p = .036$), in the US ($r = 0.28, p < .001$), and in Slovenia ($r = 0.16, p = .021$). The correlations were low, not significant, and positive in Spain ($r = 0.07, p = .311$), and negative in Italy ($r = -0.03, p = .675$).

Compared to the US, the correlation was significantly lower in Spain ($z = 2.21, p = .027$), and in Italy ($z = 3.29,$

$p = .001$). Compared to the US, the correlation was significantly higher in Germany ($z = -2.09, p = .036$). The correlations in the other countries were not significantly different from the US (all z s < 1.31).

Neuroticism

In line with our prediction, results indicated significant moderate to high negative correlations in the Netherlands ($r = -0.36, p < .001$), in the US ($r = -0.38, p < .001$), and in Germany ($r = -0.33, p < .001$). Results further indicated significant low to moderate negative correlations in Slovenia ($r = -0.28, p < .001$), in Italy ($r = -0.21, p = .002$), in the UK ($r = -0.29, p < .001$), and in Spain ($r = -0.21, p = .003$). No correlations were significantly different from the US (all z s < -1.94).

Agreeableness, Openness, and Extraversion

Results indicated no significant correlations in any country for agreeableness or openness. Contrary to our prediction, results indicated a significant low to a moderate positive correlation between the PFW and extraversion in Germany ($r = 0.20, p = .005$). Results indicated no further significant correlations with extraversion in any other country.

No correlations were significantly different from the US (for agreeableness all z s < -2.08 ; for openness all z s < -0.79 and for extraversion all z s < 1.47).

Exploratory Analyses

Due to the poor fit of the registered one-factor CFA, we conducted further exploratory analyses. Nielsen (2010) proposed a two-factor model structure (Factor 1: Subjective Financial Well-being, Items 1, 2, 3, 8; Factor 2: Objective Financial Well-being, Items 4-7 inclusive). A two-factor model would suggest that the two subscales measure different underlying constructs. Here, we found a pattern of item loading that is in line with these subscales, however, all items still rely on a subjective assessment of financial stress. Thus, a one-factor structure seems theoretically more appropriate. Yet, there does appear to be a difference between the items of the two proposed subscales. The items of the subjective subscale assess affective financial stress (e.g., feelings about the financial situation) whereas items of the objective subscale assess behavioral financial stress (e.g., living paycheck to paycheck). This suggests a multidimensionality in which all items measure subjective financial stress but may also contain further internal commonalities not captured by the one-factor structure. This could lead to measurement-related variance when using a strict one-factor model. Based on this idea, we explored a modified one-factor model which allowed for correlated errors of items assessing affective and behavioral financial stress. All

Table 5. Model fit indices for individual country CFAs for modified one-factor model of the PFWs

Country	<i>N</i>	χ^2	<i>df</i>	CFI	TLI	RMSEA	SRMR
Germany	205	22.46	8	0.99	0.96	0.09	0.03
Italy	213	21.76	8	0.99	0.95	0.09	0.02
The Netherlands	207	37.06	8	0.97	0.90	0.13	0.03
Slovenia	212	13.62	8	1.00	0.98	0.06	0.01
Spain	203	14.09	8	1.00	0.98	0.06	0.02
UK	288	11.82	8	1.00	0.99	0.04	0.01
US	222	12.55	8	1.00	0.99	0.05	0.01

Note. Results are reported rounded to 2 decimal places, however, unrounded values are used to assess model fit indices.

modifications were made concurrently. Stepwise modifications based on influential item relations were not explored because they may be dependent on the data of the current sample. The results of this modified one-factor CFA are reported in Table 5. For the modified one-factor model, the CFI was excellent (> 0.95) in all countries. The TLI was excellent (> 0.95) in all countries except Italy where it was acceptable (> 0.89). The RMSEA values are improved across all countries with the modified one-factor model. The RMSEA was excellent (< 0.06) in Slovenia, the UK, and the US, it was good (< 0.07) in Spain, questionable (< 0.09) in Italy, poor (< 0.10) in Germany, and unacceptable (> 0.10) in the Netherlands. SRMR was excellent (< 0.06) in all countries. The item loadings for the modified one-factor model are reported, in full, in Table 6.

Based on the model fit indices for the modified model, four countries were included in a multi-group CFA analysis to test the measurement invariance of the modified one-factor model across countries. The RMSEA values were unacceptably high in the Netherlands and were questionable in Germany and Italy, and thus these countries were excluded. The fit of the multi-group CFA and measurement invariance of this model was then assessed, see Table 7. Based on a cut-off of $\Delta CFI \leq 0.01$ the model achieved metric invariance. This suggests that the factor structure and loadings, but not the item functions, were similar across the four countries included in the analysis. The achieved level of invariance needs to be taken with caution, however, because they are based on a model that was defined post hoc. Additional modifications were not explored as the changes consistent with the theoretical model had already been made. Further changes to achieve partial scalar invariance would likely depend on the sample data, not the underlying model.

While overall the modifications improved the fit of the model, the effect was weaker in the three countries not included in the multi-group CFA. The “all-in” approach to model modification may have failed to appreciate the nuance of individual item functions in individual countries. Therefore, within the three countries not included in the

Table 6. Factor loading per item per country: Modified one-factor model

Item	Germany	Italy	The Netherlands	Slovenia	Spain	UK	US
Level of financial stress today	0.69	0.75	0.67	0.73	0.83	0.76	0.83
Satisfaction with present financial situation	0.76	0.80	0.73	0.73	0.78	0.76	0.82
Feelings about current financial situation	0.80	0.70	0.72	0.80	0.78	0.80	0.85
Worry about monthly living expenses	0.85	0.72	0.84	0.85	0.79	0.72	0.82
Confidence regarding financial emergency	0.66	0.71	0.67	0.70	0.74	0.77	0.77
Can't afford to go out	0.83	0.71	0.78	0.75	0.72	0.84	0.84
Paycheck to paycheck	0.81	0.78	0.75	0.83	0.82	0.85	0.85
Stress about finances in general	0.80	0.73	0.75	0.83	0.76	0.80	0.85

Note. Standardized factor loading per item of the PFW scale per country using the modified one-factor model.

Table 7. Measurement invariance for modified one-factor model

Testing level	<i>N</i>	χ^2	<i>df</i>	CFI	Δ CFI	TLI	RMSEA	SRMR
Configural Invariance	907	52.45	32	1.00		0.99	0.05	0.01
Metric Invariance	907	85.50	53	0.99	0.00	0.99	0.05	0.04
Scalar Invariance	907	181.11	74	0.98	-0.01	0.97	0.08	0.05

Note. Configural invariance can be assumed based on acceptable model fit. The change in CFI indicates that metric, but not scalar invariance is achieved as the cut-off of 0.01 is surpassed at the scalar level. Results are reported rounded to 2 decimal places, however, unrounded values are used to assess model fit indices.

multi-group CFA, we then explored the individual item functions to better understand the poorer fit of the modified model.

In Germany, differential item function analysis (DIF; see Figure 4 in Supplementary Materials, <https://osf.io/gevmr/>) indicated that Item 1 functioned differently compared with the other country samples. This may be due to a linguistic variation in the translation of that item which did not directly use the word “stress”. The translation may have created a comparatively less sensitive measurement, as flagged by the DIF which indicated that the item may not differentiate between individuals well in the German sample. Furthermore, Germany and the Netherlands showed relatively low loading for Items 1 and 5 (see Table 6). For Item 5, these differences may relate to a broad cultural difference in comfort reporting on items that reference explicit amounts of money which were present in this item only. In the Netherlands and Italy, examination of the standardized residuals showed that multiple items held stronger, or weaker, relations than predicted by the modified one-factor model (see Figure 5 in Supplementary Materials, <https://osf.io/gevmr/>).

Yet the pattern of deviations was not mirrored in each country. This further suggests that the modifications were not equally influential in all countries. However, the current modifications were theoretically coherent. Therefore, alternative modification pathways within individual countries were not explored to prevent overfitting specific country data.

Discussion

Financial stress is a global issue with clear negative consequences for health (Kahn & Pearlin, 2006) and well-being (Haushofer & Fehr, 2014). Therefore, it is crucial to have measures that accurately capture individuals' experiences of financial stress. In the current study, we aimed to test the validity and measurement invariance of the PFW scale across six European countries (Germany, Italy, the Netherlands, Slovenia, Spain, and UK) and the US. To this end, the scale was translated into five languages (German, Italian, Dutch, Slovenian, and Spanish). A one-factor structure was tested by conducting separate CFAs in each of the countries. Results were mixed, with some fit indices indicating an acceptable to the excellent fit, and others indicating a questionable to poor fit of the one-factor model. Reliability measures were excellent and overall, the scale showed convergent and discriminant validity.

Given the mixed fit indices for the one-factor structure, we did not conduct the planned multi-group CFA. Instead, we explored a modified one-factor structure. The modifications were based on the work by Nielsen (2010) who originally suggested a two-factor structure with two latent factors, namely subjective and objective financial stress. However, we decided against the two-factor structure for theoretical reasons. All items of the PFW scale rely on a subjective assessment of financial stress and thus there is no substantive support for a two-factor structure. It is also worth noting that the eigenvalue of the objective factor

did not exceed the cut-off value of 1 in the study by Nielsen (2010). At the same time, there is a measurement-related difference in the items of the two proposed subscales as items of the subjective subscale assess affective financial stress (e.g., feelings about the financial situation) whereas items of the objective subscale assess behavioral financial stress (e.g., living paycheck to paycheck). Therefore, we decided to add error correlations for the items assessing behavioral and affective financial stress respectively. Results indicated an improved fit for the modified one factor-structure in all countries. Considering the improved fit of the modified one-factor model, the PFW scale appears to be a promising tool to provide insight into the experience of financial stress within each country.

To assess if the PFW scale can be used to make comparisons between countries, we conducted a multi-group CFA testing the modified one-factor structure across Slovenia, Spain, the UK, and the US. Germany, Italy, and the Netherlands were excluded because the fit indices were still conflicted. Testing for measurement invariance, the results indicated metric invariance was achieved. This suggests an equal factor structure and equal loadings in the four countries. However, scalar and strict invariance were not reached. Considering that we had already explored a modified model, albeit, with modifications based on theoretical reasons, we did not explore further changes to achieve strict invariance based on partial invariance models. Based on the lack of scalar invariance we conclude that it is not advisable to use the scale to make comparisons between countries.

To understand why these comparisons are not tenable we investigated the factor loadings and residuals more closely. We noted that individual items may function differently in certain countries due to translational or cultural reasons. For example, Item 1 functioned differently in our German sample which may relate to linguistic differences in our translation. In turn, Item 5 functioned poorly in the Netherlands and Germany, two countries that are culturally quite similar in our sample. This difference may reflect a cultural hesitancy to respond to items relating to specific amounts of money, as required by that item. The variable factor loading of Item 5 may also relate to the study's data collection period which occurred during the COVID-19 pandemic. Item 5 relates to financial emergencies which were particularly salient during the global pandemic. Indeed, comparisons between countries in our samples may be hindered by the fact that different countries responded to the global pandemic with different financial support strategies. This may have influenced our country samples in ways we cannot account for and constrains the comparisons that can be made. Variation in our sampling proportions across countries, being either Prolific or Snowball samples, may have also influenced our results

with greater proportions of Prolific data appearing to relate to a higher likelihood of poor model fit. However, the underlying reasons for this pattern are not clear.

A noteworthy finding of the current study is the results of the discriminant validity analysis. Based on previous work (Xu et al., 2015), we predicted moderate to high correlations between PFW scores and conscientiousness and neuroticism, and low correlations for the other traits. Results were largely in line with our predictions and mostly country-invariant. However, for conscientiousness, we observed significant variation in the correlations which were high in Germany and the UK, low to moderate in the Netherlands, the US, and Slovenia, and low and nonsignificant in Spain and Italy. This suggests that in Germany and the UK in particular conscientiousness is related to financial stress. A possible explanation is that there are local structural or policy-related differences between these countries, which are affected by conscientiousness. One such difference could be the administrative burden that individuals face when they are trying to obtain financial benefits (Dubois & Ludwinek, 2015). When obtaining financial support is burdensome, higher levels of conscientiousness may aid individuals. This could have practical implications as access to financial benefits should not rely on personal administrative effort. However, further research in which personality and financial stress are assessed in a cross-cultural context is needed.

Cross-border comparisons can shed light on cultural variability within the domain of financial well-being. Previous work has shown that financial stress can be affected by cultural differences (Warmath et al., 2021). Indeed, in the current study, we found that the relationship between conscientiousness and financial stress showed significant variation between countries, with moderate to strong negative correlations in some countries and no significant correlations in others. However, when considering policy implications, understanding variability within countries may be more informative. Reducing financial stress and increasing well-being through effective policies requires a nuanced approach and interventions should be tailored to specific groups (Ruggeri et al., 2020). Therefore, using the PFW scale to explore differences within countries is a fruitful area for future research.

In conclusion, the PFW scale shows excellent reliability, and convergent and discriminant validity. The current results suggest that the scale can be used to explore specific differences within countries. However, comparisons between countries using the PFW scale are not advised. Understanding the subjective experience of financial stress has practical applications in terms of understanding and implementing effective interventions. The PFW scale provides insight into this domain and its use may benefit both research and practice on financial stress and well-being.

References

- American Psychological Association. (2021, October). *Stress in America™ 2021: Stress and Decision-making during the Pandemic*. <https://www.apa.org/news/press/releases/stress/2021/decision-making-october-2021.pdf>
- Aust, F., & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with R Markdown*. <https://github.com/crsh/papaja>
- Buabang, E. K., Ashcroft-Jones, S., Esteban Serna, C., Kastelic, K., Kveder, J., Lambertus, A., Lambertus, A., Muller, T. S., & Ruggeri, K. (2022). *Validation and measurement invariance of the Personal Financial Wellness Scale: A multinational study in 7 countries*. <https://osf.io/gevmr>
- Blanchet, T., Chancel, L., & Gethin, A. (2020). *Why is Europe more equal than the United States*. <https://halshs.archives-ouvertes.fr/halshs-03022133>
- Choi, S. W., Gibbons, L. E., with contributions from., & Crane, P. K. (2016). *Lordif: Logistic ordinal regression differential item functioning using IRT*. <https://CRAN.R-project.org/package=lordif>
- Christelis, D., Ehrmann, M., & Georgarakos, D. (2017). *Exploring differences in household debt across the United States and Euro Area countries* (No. 465). Centre for Studies in Economics; Finance (CSEF), University of Naples. <https://ideas.repec.org/p/sef/csefwp/465.html>
- Cohen, S. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan & S. Oskamp (Eds.), *The social psychology of health* (pp. 31–67). Sage Publications, Inc.
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385–396. <https://doi.org/10.2307/2136404>
- Crutzen, R., & Peters, G.-J. Y. (2017). Scale quality: Alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychology Review, 11*(3), 242–247. <https://doi.org/10.1080/17437199.2015.1124240>
- Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS One, 10*(4), Article e0121945. <https://doi.org/10.1371/journal.pone.0121945>
- Dubois, H., & Ludwinek, A. (2015). *Access to social benefits: Reducing non take up*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2806/7077>
- Eurostat. (2021). *Living conditions in Europe – Poverty and social exclusion*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Living_conditions_in_Europe_-_poverty_and_social_exclusion
- Fazlić, H., Brborović, O., Vukušić Rukavina, T., Fišter, K., Milošević, M., & Mustajbegović, J. (2012). Characteristics of people with the perceived stress in Croatia: the CroHort study. *Collegium Antropologicum, 36*(1), 165–169.
- Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology, 10*, Article 1507. <https://doi.org/10.3389/fpsyg.2019.01507>
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504–528. [https://doi.org/10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)
- Greiff, S., & Allen, M. S. (2018). *EJPA introduces registered reports as new submission format*. *European Journal of Psychological Assessment, 34*(4), 217–219. <https://doi.org/10.1027/1015-5759/a000492>
- Gutter, M., & Copur, Z. (2011). Financial behaviors and financial well-being of college students: Evidence from a national survey. *Journal of Family and Economic Issues, 32*(4), 699–714. <https://doi.org/10.1007/s10834-011-9255-2>
- Hagenaars, A. J. M., & van Praag, B. M. S. (1985). A synthesis of poverty line definitions. *Review of Income and Wealth, 31*(2), 139–154. <https://doi.org/10.1111/j.1475-4991.1985.tb00504.x>
- Harrell, F. E. Jr. (2022). *Hmisc: Harrell miscellaneous*. <https://CRAN.R-project.org/package=Hmisc>
- Haushofer, J., & Fehr, E. (2014). On the psychology of poverty. *Science, 344*(6186), 862–867. <https://doi.org/10.1126/science.1232491>
- Hofstede, G. H., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind: Intercultural cooperation and its importance for survival* (3rd ed.). McGraw-Hill.
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). *semTools: Useful tools for structural equation modeling*. <https://CRAN.R-project.org/package=semTools>
- Kahn, J. R., & Pearlin, L. I. (2006). Financial strain over the life course and health among older adults. *Journal of Health and Social Behavior, 47*(1), 17–31. <https://doi.org/10.1177/002214650604700102>
- Kamaluddin, M. R., Nasir, R., Sulaiman, W. S. W., Hafidz, S. W. M., Abdullah, J. M. A., Khairudin, R., & Zamani, Z. A. (2018). Validity and reliability of Malay version Financial Well-Being Scale among Malaysian employees. *AKADEMIKA, 88*(2), 109–120. <http://ejournal.ukm.my/akademika/issue/view/1064>
- Lee, E.-H. (2012). Review of the psychometric evidence of the perceived stress scale. *Asian Nursing Research, 6*(4), 121–127. <https://doi.org/10.1016/j.anr.2012.08.004>
- Nielsen, R. B. (2010). *Assessing financial wellness via computer-assisted telephone interviews* (SSRN Scholarly Paper No. 2061273). Social Science Research Network. <https://papers.ssrn.com/abstract=2061273>
- Olivoto, T., & L'ucio, A. D. (2020). Metan: An r package for multi-environment trial analysis. *Methods in Ecology and Evolution, 11*(6), 783–789. <https://doi.org/10.1111/2041-210X.13384>
- Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist, 16*(2), 56–69. <https://doi.org/10.17605/OSF.IO/TNRRXV>
- Peters, G.-J. Y. (2017). *Diamond plots: A tutorial to introduce a visualisation tool that facilitates interpretation and comparison of multiple sample estimates while respecting their inaccuracy*. PsyArXiv. <https://psyarxiv.com/fzh6c>
- Prawitz, A. D., Garman, E. T., Sorhaindo, B., O'Neill, B., Kim, J., & Drentea, P. (2006). Incharge Financial Distress/Financial Well-Being scale: Development, administration, and score interpretation. *Financial Counseling and Planning, 17*(1), 34–50. <https://doi.org/10.1037/t60365-000>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ravallion, M. (2012). *Poor, or just feeling poor? On using subjective data in measuring poverty*. The World Bank. <https://doi.org/10.1596/1813-9450-5968>
- Ravallion, M., & Lokshin, M. (2002). Self-rated economic welfare in Russia. *European Economic Review, 46*(8), 1453–1473. [https://doi.org/10.1016/S0014-2921\(01\)00151-9](https://doi.org/10.1016/S0014-2921(01)00151-9)
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Ruggeri, K., Garcia-Garzon, E., Maguire, Á., Matz, S., & Huppert, F. A. (2020). Well-being is more than happiness and life satisfaction: A multidimensional analysis of 21 countries. *Health and Quality of Life Outcomes, 18*(1), Article 192. <https://doi.org/10.1186/s12955-020-01423-y>
- Shapiro, G. K., & Burchell, B. J. (2012). Measuring financial anxiety. *Journal of Neuroscience, Psychology, and Economics, 5*(2), 92–103. <https://doi.org/10.1037/a0027647>

- Sommet, N., Morselli, D., & Spini, D. (2018). Income inequality affects the psychological health of only the people facing scarcity. *Psychological Science*, 29(12), 1911–1921. <https://doi.org/10.1177/0956797618798620>
- Tierney, N. (2017). Visdat: Visualising whole data frames. *Journal of Open Source Software*, 2(16), Article 355. <https://doi.org/10.21105/joss.00355>
- Vallejo, M. A., Vallejo-Slocker, L., Fernández-Abascal, E. G., & Mañanes, G. (2018). Determining factors for stress perception assessed with the Perceived Stress Scale (PSS-4) in Spanish and other European samples. *Frontiers in Psychology*, 9, Article 37. <https://doi.org/10.3389/fpsyg.2018.00037>
- Warmath, D., Chen, P.-J., Grable, J., & Kwak, E. J. (2021). Soft landings: Extending the cushion hypothesis to financial well-being in collectivistic cultures. *Journal of Consumer Affairs*, 55(4), 1563–1590. <https://doi.org/10.1111/joca.12408>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), Article 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., François, R., Henry, L., & Müller, K. (2022). *Dplyr: A grammar of data manipulation*. <https://CRAN.R-project.org/package=dplyr>
- Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Chapman & Hall/CRC. <https://yihui.org/knitr/>
- Xu, Y., Beller, A. H., Roberts, B. W., & Brown, J. R. (2015). Personality and young adult financial distress. *Journal of Economic Psychology*, 51, 90–100. <https://doi.org/10.1016/j.joep.2015.08.010>
- Zhu, H. (2021). *kableExtra: Construct complex table with "kable" and pipe syntax*. <https://CRAN.R-project.org/package=kableExtra>

History

Received June 24, 2019
 Revision received October 23, 2022
 Accepted October 24, 2022
 Published online December 8, 2022
 EJPA Section / Category Miscellaneous / Other

Authorship

Eike K. Buabang, conceptualization, supervision, writing – original draft, investigation, data curation, formal analysis, visualization,

writing – review & editing; Sarah Ashcroft-Jones, writing – original draft, investigation, formal analysis, visualization, writing – review & editing; Celia Esteban Serna, writing – original draft, investigation, resources; Katarina Kastelic, writing – original draft, investigation, resources, writing – review & editing; Jakob Kveder, writing – original draft, investigation, resources; Amanda Lambertus, writing – original draft, investigation; Tasja S. Müller, writing – original draft, investigation, resources; Kai Ruggeri, conceptualization, funding acquisition, supervision, writing – review & editing.

Open Science

We report how we determined our sample size, all data exclusions, all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models. If we use inferential tests, we report exact *p* values, effect sizes, and 95% confidence or credible intervals.

Open Data: I confirm that there is sufficient information for an independent researcher to reproduce all of the reported results (Buabang et al., 2022).

Open Materials: I confirm that there is sufficient information for an independent researcher to reproduce all of the reported methodology (Buabang et al., 2022).

Preregistration of Studies and Analysis Plans: This study was preregistered with an analysis plan (Buabang et al., 2022).

The Stage-1 report, materials, data, code, and supplementary materials are available on the Open Science Framework: <https://osf.io/gevmr/>.

Funding

Funding for data collection was provided through Columbia University Mailman School of Public Health, Department of Health Policy and Management.

ORCID

Eike K. Buabang
 <https://orcid.org/0000-0002-3057-0819>

Eike K. Buabang

Trinity College Institute of Neuroscience
 Trinity College Dublin
 Dublin 2
 Ireland
eike.buabang@tcd.ie