Journal of Affective Disorders 387 (2025) 119524

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Six-country psychometric comparison of women responses to the Depression Anxiety Stress Scale (DASS) and the Child and Adolescent Behaviour Inventory (CABI) across cultures and time

Francisco Antonio Calderón Alfaro [a,*], Jamie Lachman [a,b], Catherine L. Ward [b], Qing Han [a], Rosanne Jocson [c], Ivo Kunovski [d,e], Stephanie Eagling-Peche [a], Rumaya Juhari [f], Kufre Okop [b], Jennel Reyes [g], Viorel Babii [h], Lucie Cluver [a,b], Liane Peña Alampay [g], Marija Raleva [e], Frances Gardner [a]

[a] University of Oxford, United Kingdom
[b] University of Cape Town, South Africa
[c] Nanyang Technological University, Singapore
[d] Institute Alternativa, North Macedonia
[e] Ss. Cyril and Methodius University Skopje, North Macedonia
[f] University of Putra Malaysia, Malaysia
[g] Ateneo de Manila University, Philippines
[h] Health for Youth Association, Republic of Moldova

ARTICLE INFO

Keywords:
Depression
Externalizing behaviour
DASS
CABI
Cross-cultural comparisons
Psychometrics

ABSTRACT

Background: The Depression Anxiety Stress Scale (DASS) and the Child and Adolescent Behaviour Inventory (CABI) are freely accessible tools used to assess depression and externalizing symptoms, respectively. There is limited psychometric evidence on how these scales hold over time and across cultures. This study aims to evaluate the internal structure of both scales and their invariance across six countries and over two timepoints.
Methods: Data from the ParentChat Pilot Study included information from 566 adults (85.51 % female) with children aged 2 to 17. Only female data from North Macedonia, Malaysia, Moldova, Montenegro, Philippines, and South Africa were used. Confirmatory factor analysis was employed to assess invariance of each construct.
Results: At baseline, the DASS depression subscale achieved scalar invariance across five countries (CFI = 0.973, TLI = 0.956, RMSEA = 0.060, SRMR = 0.069) and across timepoints (CFI = 0.990, TLI = 0.979, RMSE = 0.037, SRMR = 0.034). The CABI extract achieved scalar invariance across two countries (CFI = 0.968, TLI = 0.963, RMSEA = 0.068, SRMR = 0.085) and across timepoints (CFI = 0.983, TLI = 0.980, RMSEA = 0.044, SRMR = 0.067).
Limitations: Small sample sizes, disproportionate female sample, and use of subscales or item extracts may limit generalizability. The scales were also not validated for all age groups used.
Conclusions: The study provides evidence of validity for the internal structure of the DASS and CABI extracts across cultures and over time. These findings support the use of these open-access tools in resource-limited settings to promote local research.

1. Introduction

Research on the global burden of disease has revealed a consistent increase in life expectancy since the 1990s. As populations continue to age and healthcare services become more accessible, there is increasing interest on how these extra years are lived, with emphasis on mental health and overall quality of life. Since 2007 depressive disorders became the third biggest cause of years lived with disability (YLD) across all ages and sexes, a value that has only increased from 2007 to 2017 and now accounts for 14 % of all YLD (James et al., 2018). For children and adolescents, a meta-analysis of 27 different countries estimated that 13.4 % are actively affected by any mental disorder, which could rise up

* Corresponding author at: Barnett House, 32 -37 Wellington Square, Oxford OX1 2ER, United Kingdom.
E-mail address: Francisco.calderon@spi.ox.ac.uk (F.A. Calderón Alfaro).

https://doi.org/10.1016/j.jad.2025.119524
Received 9 October 2024; Received in revised form 13 May 2025; Accepted 26 May 2025
Available online 27 May 2025

to 17.4 % if symptoms, rather than a diagnosable disorder are considered (Polanczyk et al., 2015). It is estimated that 241 million youths worldwide are affected by a mental disorder. The most common mental health disorders are anxiety disorders (117 million), disruptive behaviour disorder (113 million), ADHD (63 million), and depressive disorders (47 million) (Polanczyk et al., 2015). More recent studies have identified higher rates for some types of mental health disorders, possibly linked to the onset of the COVID-19 pandemic (Racine et al., 2021).

To estimate the prevalence of a mental health disorder, adequate measuring tools are needed. This generates a series of challenges by requiring psychological scales to be translated, contextualized, adapted, and tested (Perneger et al., 1999). Psychometric studies in high income countries are often the most cited reflecting the greater level of resources to conduct this type of research locally (Cai et al., 2023; Polanczyk et al., 2015; Racine et al., 2021). This, however, further widens the knowledge gap with low- and middle-income countries (LMIC) (Kaku et al., 2022; Limenih et al., 2024). The process of adapting psychometric scales in different contexts can result in more credible data as the scales are contextualized and can be optimized or shortened. Moreover, showing evidence of the validity of the measures across different contexts indicates that the results are comparable across regions (Gugerty and Karlan, 2018). Adaptations and evidence of validity provide information for other researchers using the scales to make contextualized informed decisions.

The Depression Anxiety Stress Scale (DASS) and the Child and Adolescent Behaviour Inventory (CABI) are freely available measuring instruments and are thus affordable in LMIC. The DASS-21 was initially developed and cross-validated between clinical groups and community samples (Antony et al., 1998; Ng et al., 2007), although normative data has been drawn for population estimates (Henry and Crawford, 2005), and its factor structure has been evaluated before (Szabó, 2010). Additionally, it has been extensively adapted and tested in cultures such as Hindi (Singh et al., 2013), Arabic (Moussa et al., 2017), Chinese (Jiang et al., 2020), and Italian (Bottesi et al., 2015), amongst others (Murphy et al., 2024). On the other hand, the CABI is a more recent measurement tool assessing a wide range of disorders, including child internalizing and externalizing behaviour (Cianchetti et al., 2013). The CABI has been used on children aged 6–17 (Cianchetti et al., 2013, 2017) and predictive validity for clinical populations has been tested (Cianchetti et al., 2020). It has also been adapted for use Brazil (Costa et al., 2023) and Italy (Cianchetti et al., 2020). Although within country studies have been conducted for both scales to a different extent, there is a void of research focusing on comparing psychometric properties across cultures and stability of the constructs of depression and child behavioural symptoms as measured by these scales across time. This puts into question the comparability of the constructs between countries and between different timepoints (American Educational Research Association et al., 2014).

This study makes use of secondary data collected for the DASS and CABI from six countries – North Macedonia, Malaysia, Moldova, Montenegro, Philippines, and South Africa – at several timepoints. This presents an exceptional opportunity to compare the internal structure of these scales within the different countries, between the countries and across timepoints, providing evidence of validity for cross-cultural and cross-time comparisons (American Educational Research Association et al., 2014; Boateng et al., 2018; Putnick and Bornstein, 2016). The aims of this study are therefore to: a) test the internal structure of the DASS and CABI within each country at baseline, b) compare the factor structure of the scales across countries, and c) compare the factor structure of the scales of the combined countries at different timepoints.

2. Method

This study makes use of data from an intervention study named ePLH Pilot Study: Online Support Parent Groups - ParentChat. With the aim to improve positive parenting and reduce child maltreatment, the original study provided a 8-session parenting intervention delivered via messaging apps (e.g., WhatsApp) to adult caregivers during the COVID-19 pandemic (ClinicalTrials.gov ID: NCT04809272). The online intervention is an adaptation of the in-person Parenting for Lifelong Health programmes for parents of children aged 2–17 years, that were originally developed and tested in South Africa (Cluver et al., 2016; Lachman et al., 2017; Ward et al., 2020).

2.1. Participants

Participants were recruited with the aid of several partnered organisations. Data collection was conducted online and led by the institutions within the countries. For North Macedonia, Institute Alternativa led recruitment; in Malaysia it was the National Population and Family Development Board; in Moldova, the Health for Youth Association was responsible; Montenegro had a collaboration of individuals from health centres, kindergartens, and NGOs; in the Philippines the study partnered with the Valenzuela City Social Welfare and Development Office; lastly in South Africa the organisation Clowns Without Borders South Africa was responsible for recruitment. There was no intentional selective recruitment for families. It may be the case that given the areas and populations the partnered organisations work with, families could be prone to having children and adolescents with higher externalizing problems and caregivers with higher depression symptoms.

At baseline, 566 adult participants consented to participation and completed the first questionnaire across all countries. Participant ages ranged from 18 to 71 ($M = 39.58$, $SD = 8.59$) of which 31 were male (5.48 %), 484 were female (85.51 %), and 51 did not disclose their sex (9.01 %). Due to small number of men, only women were considered for analysis. All women were required to respond thinking of a single child (even if they had multiple) whose age was 2 to 17 years ($M = 10.96$, $SD = 4.35$); of those 254 were boys (44.88 %), 268 were girls (47.35 %), and 44 (7.88 %) did not have their sex disclosed. Most women reported being the biological parent of the child ($n = 400$), others mentioned being a stepparent ($n = 1$), grandparent ($n = 24$), an aunt ($n = 23$), a sibling ($n = 16$), a cousin ($n = 2$), foster parent ($n = 14$), or something different ($n = 4$). Since data on the DASS and the CABI were analysed independently, complete case analysis resulted in slightly different sample sizes (Fig. 1).

2.2. Measures

Data was originally collected at baseline for the intervention and ten weeks afterwards. The first timepoint was before the intervention was delivered with the last measurement point being after the intervention delivery had concluded. These measurements were used as screening tools to assess outcomes associated with the delivery of the intervention in this case, parent depression and child externalizing symptoms.

The depression subscale of the Depression Anxiety Stress Scale (DASS-D, 7 items) was used to measure parental depression symptoms (Antony et al., 1998). Item response options ranged from 0 (*Never*) to 3 (*Always*) with minimum score of 0 and maximum score of 21, and higher scores indicating more depression symptoms (Supplementary Table 1).

Fourteen items from the Child and Adolescent Behaviour Inventory were used to measure child externalizing problems (Cianchetti et al., 2013). An extract of items encompassed irritability, externalizing behaviour of oppositional-defiant disorder, and conduct disorder were used in original study to capture children externalizing problems. Items contain three response options ranging from 0 (*Not True*) to 2 (*Very True*) with a minimum score of 0 and a maximum score of 42, with higher scores indicating more externalizing problems (Supplementary Table 2).

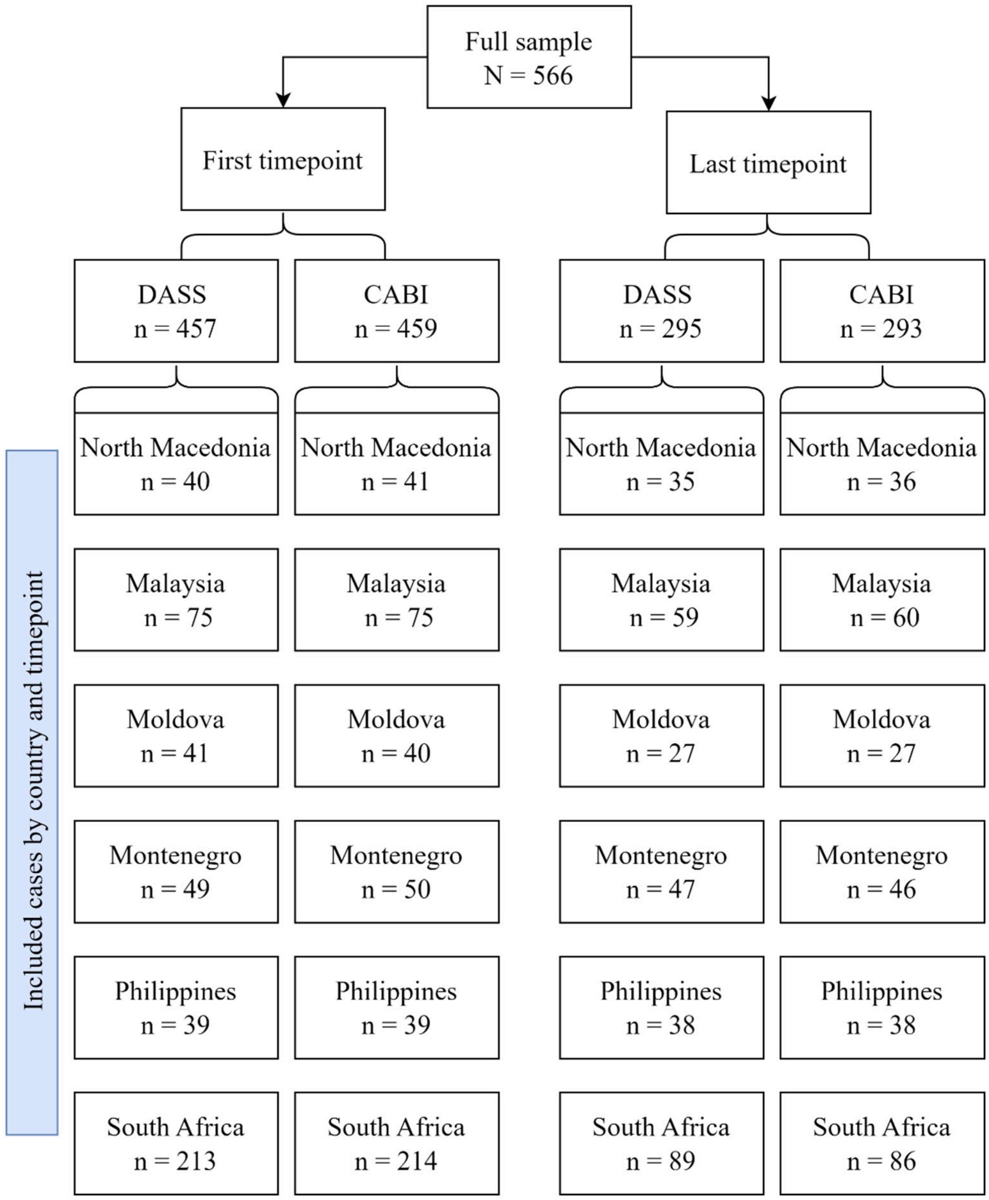

Fig. 1. Flow-chart selection of participants in the analytical sample.

2.3. Analysis plan

Analyses were carried out in R version 4.2.1 (R Project, 2023) paired with RStudio Desktop version 2022.02.3 + 492 (Rstudio, 2022), primarily using the lavaan package (Rosseel, 2012). Response option patterns for each scale were reviewed for missingness of response options for each country and timepoint. Using the first timepoint of all countries, correlation matrixes were computed to inspect individual item correlations. Similarly, McDonald's Omega and Cronbach's Alpha were used to analyse both subscales to compare if values were consistent with original scales (Antony et al., 1998; Cianchetti et al., 2013). Confirmatory factor analysis (CFA) was used to test the single factor structure of child externalizing problems from the CABI and adult depression symptoms from the DASS. In the case of the CABI, only the 11 items that loaded onto the same factor from its original study were considered as part of the factor analysis, this excluded CABI12, CABI13, and CABI14 (Cianchetti et al., 2013). For the CFAs, given the response option and data patterns, the weighted least squares means and variance adjusted (WLSMV) estimator was used for the CABI as used in other psychometric papers evaluating the measurement (Costa et al., 2023) with its factor indicators treated as ordered categorical (Flora and Curran, 2004). Since items in the DASS scale contain more response options, both the WLSMV (Bottesi et al., 2015; Jiang et al., 2020) and the Maximum Likelihood Robust (MLR) (Henry and Crawford, 2005) have been previously used. For the DASS-D, we used the WLSMV estimator initially, and if models did not fit properly, Maximum Likelihood Robust (MLR) was used for all models with indicators being considered in a continuous scale (Little, 2023).

To show proof of validity for the internal structure CFA analyses were conducted within each of the six countries at the first timepoint as it had the most completion (American Educational Research Association et al., 2014). CFA models were considered to fit well if root mean square error of approximation (RMSEA) and standardized root mean squared residual (SRMR) were <0.05 (MacCallum et al., 1996; Schermelleh-Engel et al., 2003) and Tucker-Lewis index (TLI) and CFI were >0.95 (Hu and Bentler, 1999). Models were acceptable if CFI and TLI were >0.90 and RMSEA and SRMR were <0.08 (Little, 2023). If models did not reach acceptable fit measures, modification indices were checked and theoretically plausible changes that had the highest modification index value were considered for the specific model (Dimitrov, 2010).

Modifications made to individual models were used to inform the model used for invariance testing. A decrease in the comparative fit index (CFI) of >0.01 was considered to indicate non-invariance (Boateng et al., 2018; Little, 2023; Putnick and Bornstein, 2016).

Invariance testing was conducted for both scales at two different levels. The initial level compared individual countries at the first timepoint. After confirming invariance between countries, then the regions were pooled at the first and last timepoints and compared. Any additions to the model needed to achieve adequate fit measures during the first timepoint invariance testing were carried over to the timepoint comparison invariance testing. On the second stage, individual models were fit for each timepoint and subsequently compared through the invariance testing. Invariance testing followed the traditional comparisons between Configural model, CFA with each group estimated differently; Metric model, a model with equal loadings between groups; and Scalar, a model with equal loadings and thresholds between groups. In the case that models did not achieve satisfactory fit measures, partial invariance was explored. Selection of parameters to be freed up would the ones with the highest chi-square value which were also significant ($p < .05$), one at a time (Brown et al., 2017; Dimitrov, 2010).

3. Results

Response patterns for both the first and last timepoint for all countries can be observed in Supplementary Tables 3–4 for the DASS-D and Supplementary Tables 5–6 for the CABI. Missing response options between countries were inspected both at the first and last timepoints in the case that the WLSMV estimator was used instead of MLR. For the DASS-D, comparisons between countries would not have been possible using the WLSMV estimator due to missing different response options. For the CABI, only South Africa and Malaysia had full response options that could enable between country comparisons with said estimator.

The correlation matrix for DASS-D items in all countries for the first timepoint can be seen in Supplementary Table 7; the CABI correlation matrix can be found in Supplementary Table 8. For the DASS-D, correlations are mostly stable, with DASS6 and DASS7 having the highest correlation ($r = 0.628$), and DASS2 and DASS6 having the lowest value ($r = 0.284$). The CABI results showed some low correlations with CABI7 and CABI11 having the lowest value ($r = 0.064$) and CABI5 with CABI6 having the highest value ($r = 0.606$). Initial internal consistency checks were conducted for the DASS-D using both Cronbach's Alpha (0.84) and MacDonald's Omega (0.88). For the CABI, reliability values were high as well for Cronbach's Alpha (0.81) and MacDonald's Omega (0.85).

When conducting the CFA, the WLSMV estimator was initially tested for the DASS-D, however, model fits were extremely low or failed to compute for individual countries. The MLR estimator was used for all the models fitted for the DASS-D. For the CABI CFA the version with 11-items was considered as the remaining three items (CABI12, CABI13, CABI14) had not loaded into any factor in the original paper (Cianchetti et al., 2013). This resulted in two models, an 11-item single factor structure, and a 11-item two factor structure. Likely due to small sample size, models following the two-factor structure resulted in multiple errors when estimated (Supplementary Table 9). Because of this, the one factor structure for general children externalizing problems was used.

One-factor CFA models were fitted for the 7-item DASS-D and 11-item CABI within each country for the first timepoint (Table 1). Initial result for individual models did not reach the acceptable thresholds for most models, and so modification indices were checked, and iterative additions were included in the model until a satisfactory composition was reached. For the DASS in North Macedonia, covariances were added between DASS5 with DASS6 and DASS4 with DASS6. For Malaysia, covariances were added between DASS6 with DASS7, DASS5 with DASS7, and DASS2 with DASS4. For Moldova, no covariances were added as the model resulted in perfect fit measures, which may be a sign that the model is overfitted for the data provided. For Montenegro, a covariance between DASS3 with DASS5 was added. For Philippines, a covariance between DASS3 with DASS5 was added. Lastly, for South Africa, a covariance between DASS6 with DASS7 was added.

Table 1
Fit measures of CFA at first timepoint by country for DASS-D and CABI.

Country	CFI	TLI	RMSEA	SRMR
DASS-D				
North Macedonia	0.992	0.993	0.042	0.053
Malaysia	0.987	0.975	0.050	0.056
Moldova	1.000	1.000	0.000	0.043
Montenegro	0.980	0.968	0.053	0.070
Philippines	0.944	0.909	0.077	0.076
South Africa	0.963	0.939	0.058	0.049
CABI				
North Macedonia	0.984	0.977	0.046	0.124
Malaysia	0.983	0.976	0.068	0.106
Moldova	0.983	0.979	0.050	0.153
Montenegro	0.988	0.983	0.038	0.184
Philippines	0.968	0.958	0.072	0.144
South Africa	0.958	0.942	0.076	0.085

For the CABI in North Macedonia, covariances were added between CABI1 with CABI5, CABI3 with CABI6, CABI9 with CABI10 and CABI11, CABI10 with CABI11. For Malaysia covariances were added between CABI5 with CABI6, CABI9 with CABI10, CABI11, and CABI8, and CABI10 with CABI11. For Moldova covariances were added between CABI2 with CABI4. For Montenegro covariances were added between CABI3 with CABI4, CABI9 with CABI11, and CABI5 with CABI6. For Philippines covariances were added between CABI6 with CABI7 and CABI3 with CABI10. Lastly, for South Africa covariances were added between CABI5 with CABI6, CABI6 with CABI7, CABI9 with CABI11, and CABI10 with CABI11 All additions to the models were done one by one in the order mentioned. Further additions to the models did not significantly improve the models or made no theoretical sense.

Due to the resulting fit measures for the CFAs per country for the first timepoint, data from all countries except Moldova were combined to test invariance between countries for the DASS-D; only Malaysia and South Africa were combined for the CABI due to having responses for all categories in its items and enabling the usage of the WLSMV estimator. For the DASS-D, the configural model contained the covariances previously added for the CFA models within the 5 countries and obtained satisfactory fit measures (Table 2). When loadings were constrained to be equal for Metric invariance, an additional covariate was added between DASS4 with DASS7. Scalar invariance was achieved after adjusting for partial invariance sequentially, first the intercept for DASS2 was allowed to vary between groups, next the DASS4 was released of constrains, and lastly DASS5 intercept was allowed to vary (first and last change were prompted by differences in Montenegro, second change was prompted by differences in South Africa). Results suggest metric invariance of the measurement model between the five countries, and partial scalar invariance between them.

Table 2
Invariance testing for between countries for DASS-D and CABI.

Model	CFI	TLI	RMSEA	SRMR
DASS-D				
Configural	0.982	0.947	0.066	0.037
Metric	0.979	0.959	0.058	0.069
Scalar	0.973	0.956	0.060	0.069
CABI				
Configural	0.972	0.960	0.071	0.086
Metric	0.970	0.961	0.070	0.095
Scalar	0.968	0.963	0.068	0.085

Note: comparisons were done for the first timepoint. For the DASS-D, only North Macedonia, Malaysia, Montenegro, Philippines, and South Africa were compared. For the CABI, only Malaysia and South Africa were compared.

For the CABI, the configural model contained the covariances previously added for the CFA models within both Malaysia and South Africa. Configural model obtained satisfactory fit measures (Table 2), however, when loadings were constrained to be equal for Metric invariance, an additional covariate was added between CABI2 with CABI4. Scalar invariance was achieved afterwards without any further changes to the model. Results suggest invariance of the measurement model between Malaysia and South Africa at the first timepoint.

After achieving invariance between the pooled countries, we tested invariance across time between the baseline and last timepoint (Table 3). Because of the achieved invariance at the country level, the five countries were pooled and compared between timepoints for the DASS-D and only South Africa and Malaysia for the CABI. The resulting model structure from the between country comparisons was used for both timepoints. The DASS-D model for the last timepoint required the addition of one additional covariance between DASS1 with DASS5. The CABI model for both timepoints did not require any changes.

Results of invariance testing across the first and last timepoint show positive results (Table 4). For the DASS-D, invariance was achieved at the Configural and Metric level without further additions to the model. Scalar invariance was achieved after adjusting for partial invariance. First, DASS1 intercept was released, then DASS2 intercept was released. No further changes were included in the model to achieve Scalar invariance. For the CABI, invariance was achieved at the Configural, Metric and Scalar level without further additions to the models. It should be noted that when pooling data, the item CABI1 for Malaysia at the last timepoint does not possess any responses in one of the options which lead to errors when estimating the models. Resulting loadings for the country pooled model (only one timepoint) and the country-timepoint (two timepoints) pooled models can be observed in Table 5.

4. Discussion

Scores from psychometric measures are often compared between countries and between different timepoints, often with little evidence whether these constructs are stable between the compared groups. This study set out to evaluate the internal structure of the freely available DASS and CABI measurement tools in multiple countries. It examined the internal structure of the DASS-D and CABI within and between countries and between different timepoints using secondary data from an all female sample collected in North Macedonia, Malaysia, Moldova, Montenegro, Philippines, and South Africa. The statistical analyses showed promising results for the internal structure of the DASS-D and CABI within countries, which aligns with previous studies on the psychometric properties of the measures (Bottesi et al., 2015; Costa et al., 2023; Jiang et al., 2020; Moussa et al., 2017; Singh et al., 2013; Szabó, 2010). This study makes an additional contribution by examining between country and time comparisons, which are generally more limited in literature, and not currently available for the DASS and CABI. Cross country and time comparisons provide information on how comparable the measurements are. The results presented here provide evidence of construct robustness for the freely available DASS and CABI which should aid researchers make better informed comparisons of the results of these scales.

Table 3
Fit measures of CFA for first and last timepoint for the pooled countries for DASS-D and CABI.

Timepoint	CFI	TLI	RMSEA	SRMR
DASS-D				
First timepoint	0.995	0.983	0.033	0.043
Last timepoint	0.990	0.958	0.052	0.023
CABI				
First timepoint	0.974	0.961	0.063	0.065
Last timepoint	0.978	0.967	0.047	0.063

Note: For the DASS-D, only North Macedonia, Malaysia, Montenegro, Philippines, and South Africa were pooled. For the CABI, only Malaysia and South Africa were pooled.

Table 4
Invariance testing between first and last timepoint for DASS-D and CABI.

Model	CFI	TLI	RMSEA	SRMR
DASS-D				
Configural	0.998	0.991	0.025	0.017
Metric	0.996	0.989	0.027	0.030
Scalar	0.990	0.979	0.037	0.034
CABI				
Configural	0.983	0.975	0.048	0.060
Metric	0.979	0.972	0.051	0.074
Scalar	0.983	0.980	0.044	0.067

Note: For the DASS-D, only North Macedonia, Malaysia, Montenegro, Philippines, and South Africa were pooled. For the CABI, only Malaysia and South Africa were pooled.

Table 5
Loadings of items for resulting CFA without grouping by country for 7-item DASS-D and 11-item CABI.

Item	Loading	Standard error	*P*-value
Country pooled			
DASS1	1.000	–	–
DASS2	0.974	0.118	<0.001
DASS3	1.115	0.112	<0.001
DASS4	1.142	0.119	<0.001
DASS5	1.123	0.125	<0.001
DASS6	0.815	0.126	<0.001
DASS7	0.839	0.107	<0.001
Country-timepoint pooled			
DASS1	1.000	–	–
DASS2	0.899	0.084	<0.001
DASS3	1.002	0.078	<0.001
DASS4	1.037	0.086	<0.001
DASS5	1.047	0.088	<0.001
DASS6	0.825	0.090	<0.001
DASS7	0.830	0.079	<0.001
Country pooled			
CABI1	1.000	–	–
CABI2	1.005	0.114	<0.001
CABI3	0.902	0.109	<0.001
CABI4	1.016	0.107	<0.001
CABI5	0.912	0.110	<0.001
CABI6	0.802	0.100	<0.001
CABI7	0.696	0.112	<0.001
CABI8	1.012	0.104	<0.001
CABI9	0.985	0.124	<0.001
CABI10	0.763	0.119	<0.001
CABI11	0.867	0.142	<0.001
Country-timepoint pooled			
CABI1	1.000	–	–
CABI2	1.015	0.092	<0.001
CABI3	0.913	0.087	<0.001
CABI4	1.035	0.088	<0.001
CABI5	0.969	0.088	<0.001
CABI6	0.796	0.082	<0.001
CABI7	0.666	0.091	<0.001
CABI8	0.980	0.085	<0.001
CABI9	0.880	0.099	<0.001
CABI10	0.794	0.092	<0.001
CABI11	0.213	0.036	<0.001

Note: For the DASS, only North Macedonia, Malaysia, Montenegro, Philippines, and South Africa were pooled. For the CABI, only Malaysia and South Africa were pooled.

Given the smaller sample sizes examined within each country, decisions had to be taken regarding the estimator and the factor structure of the scales. Differences in the literature for the DASS regarding the usage of WLSMV (Bottesi et al., 2015; Jiang et al., 2020) or MLR (Henry and Crawford, 2005) estimator led to both being tested. The MLR estimator resulted in a better estimator for the current model. As the DASS contains four response options, it is more flexible on the type of estimators that can be applied (Little, 2023). On the other hand, it could be that given to small sample size per grouping category, not all response options were available and thus it resulted in poorer fit measures. For the CABI, only the WLSMV estimator was used, which fit the models adequately as it has done in previous studies (Costa et al., 2023). However, given only a fraction of the full CABI was used during the original study, paired with having small sample size per grouping category, it became difficult to replicate factors tested in more recent studies (Cianchetti et al., 2020; Costa et al., 2023), and factor structure was then decided based on the original scale results (Cianchetti et al., 2013).

Internal consistency checks for both the DASS-D and CABI showed good Cronbach's Alpha and MacDonald's Omega. Values for the DASS-D were high although slightly lower than those obtained in the original inception of the scale (Antony et al., 1998); reliability values for the CABI were high as very similar to those obtained in the original paper (Cianchetti et al., 2013). Correlation matrixes for both the DASS-D and CABI show mostly moderate correlations, with the CABI having a few more low correlations. The original CABI scale (Cianchetti et al., 2013) references externalizing problems through four different factor structures, a core set of externalizing problems (CABI5 to CABI11), irritability (CABI1 to CABI4), three added items that were considered clinically relevant (CABI12, CABI13, CABI14), and everything added as a factor. This would explain differences between individual item correlations.

CFA within each country at baseline assessments showed positive results for the internal structure of the depression subscale of the DASS-D. Results for North Macedonia, Malaysia, Montenegro, and South Africa show satisfactory results supporting the notion of the internal structure working within each country similar to previous studies (Bottesi et al., 2015; Henry and Crawford, 2005; Jiang et al., 2020; Murphy et al., 2024). Results from Philippines are borderline acceptable, although it is important to note that it also has the smallest sample size which can make fit indices less stable (Little, 2023; Wolf et al., 2013). On the other hand, functionality of the scale within Moldova couldn't be tested as the model produced almost perfect fit measures. Overall, these results show proof of validity for the internal structure functionality within North Macedonia, Malaysia, Montenegro, Philippines, and South Africa.

Likewise, CFA within each country for the CABI show favourable results for the CFI and TLI estimators with RMSEA mostly acceptable and SRMR being high for all countries. One explanation for this could be the small sample size as both RMSEA and SRMR tend to be more susceptible to sample sizes, however the likely explanation for these is the lower correlations seen in the correlation matrix (Little, 2023; MacCallum et al., 1996; Schermelleh-Engel et al., 2003), as opposed to the DASS-D. Generally, in terms of internal structure, results from the factorial analysis of the CABI for each individual country should be interpreted with caution. Most fit measures show acceptable or excellent results, moreover the underlying issue likely with the covariance matrix may indicate that the items are not working as well as intended for a single factor structure. In terms of psychometric validity, these should be weighted together with other forms of validity (e.g. content validity) that can further inform the composition of the structure; similar to how the original scale included clinically relevant items into the factor structure without statistical evidence (Cianchetti et al., 2013).

The diversity of countries analysed for invariance makes a significant contribution to literature as between culture comparisons are uncommon, but results are assumed to be comparable (e.g. comparing DASS depression scores from one paper to another). For between country invariance testing of the DASS-D, we were able to compare data across North Macedonia, Malaysia, Montenegro, Philippines, and South Africa. Configural and Metric invariance were achieved without many modifications, with Scalar invariance requiring partial invariance to be included into the model for some parameters of the model. Guidelines for acceptable partial invariance are inconsistent with some suggesting at least no changes on a majority of items, half of items (Putnick and Bornstein, 2016), or a proportion of the overall model parameters (Dimitrov, 2010). Regardless of the criteria, the given partial invariance results from the study would at worst be borderline (e.g. the number of model parameters) but within the acceptable range. This result is promising given the high number of groups tested which would be expected to introduce more noise. For the CABI, only Malaysia and South Africa were compared since the lack of answered response option categories by the participants in other countries would have resulted in several model errors. Invariance testing for these two countries somewhat resemble the results obtained in the within country cases with slight improvements: CFI and TLI were high and acceptable, RMSEA was just acceptable and SRMR was just above the acceptable but marginally lower than the within country comparison. This is likely explained by the pooling of the two countries and greater sample sizes which made the covariance matrix more stable (Little, 2023; MacCallum et al., 1996; Schermelleh-Engel et al., 2003); additionally, both countries individually had the lowest SRMR values from all of them. These results provide evidence for the comparability of both the DASS-D and CABI scales between the respective countries included in each comparison (Boateng et al., 2018; Brown et al., 2017; Dimitrov, 2010; Little, 2023; Putnick and Bornstein, 2016). As open-access tools, this has broad implications for as it provides support for the widespread use of these open access tolls across cultures and settings, without further costly adaptation.

After achieving Scalar invariance between countries, the same countries were pooled and tested for time invariance. This approach was favoured given that an intervention had taken place in between the measurements which would have likely influenced the raw scores or comparisons, nonetheless the loadings of the items should remain stable across time if invariant. Individual timepoint models for both the DASS-D and CABI reached satisfactory fit measures in a straightforward manner. This should provide additional evidence of the invariance between countries within both timepoints, as well of the added stability to the models due to the increased sample size. Invariance testing between timepoints was similar to the testing between countries but resulted in fewer model adaptations. The DASS-D reached Configural, Metric, and partial Scalar invariance with fewer adaptations. The CABI reached Configural, Metric, and Scalar invariance without any further additions to the models. This provides evidence for the comparability of the constructs across time-cultures, meaning the constructs can be expected to be measuring the same thing between North Macedonia, Malaysia, Montenegro, Philippines, and Malaysia for the DASS-D, and Malaysia and South Africa for the CABI at different timepoints (Boateng et al., 2018; Little, 2023; Putnick and Bornstein, 2016).

The opportunity of such conducting research on such a culturally diverse sample came with limitations. The small sample sizes of the initial study per country limited the testing of these measures for Moldova and invariance testing of the CABI due to lack of categorical response options chosen by participants in the different countries. At the same time this resulted in using MLR instead of WLSMV as the estimator for the DASS-D. Because of the smaller sample within countries the decision to keep only women was taken as men would be spread out thinly across the different countries. The usage of secondary data also resulted in subscales being used instead of complete scales. Similarly, both of this resulted in the usage of the CABI on children from diverse developmental stages, some toddlers (reported by parents) being the more problematic as some of the items may not be adequately adapted. Also influencing all of these points is the fact that these data were collected as part of an intervention study which had the intention of influencing both outcomes from the scales included in this study, being parental depression and child externalizing symptoms. This resulted in less variation in the items for the last timepoint. On a different note, even though the scales were contextualized to every country, more data could have been collected to show evidence of validity from different perspectives (e.g. cognitive interviews to understand if participants were interpreting the questions similarly across the different countries, or selecting populations that had already been clinically assessed).

5. Conclusions

This study provides evidence of validity for the internal structure of the DASS-D and CABI. Testing of both scales for individual countries resulted positively for the internal structure for the DASS-D subscale and borderline acceptable for the CABI subscale. As the most novel part the study, invariance at the Scalar level was achieved for the five-country comparison of the DASS-D and the two-country comparison of the CABI. Similarly, time invariance was also achieved between the same countries for the two different timepoints of the DASS-D and CABI. This study provides evidence for the robustness of both scales being comparable across cultures and time. Our findings suggest that these open-access tools are widely applicable. Their use would contribute to decolonising research through using valid, reliable alternatives to for-profit instruments developed in high-income countries. Future studies should investigate the functionality of the full scales with bigger sample sizes, specially including more males; explore the functionality of the CABI externalizing subscale as a two-factor structure; and explore other proofs of validity of the scales in different countries.

CRediT authorship contribution statement

Francisco Antonio Calderón Alfaro: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jamie Lachman:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Catherine L. Ward:** Writing – review & editing, Project administration. **Qing Han:** Writing – review & editing, Investigation. **Rosanne Jocson:** Writing – review & editing, Project administration. **Ivo Kunovski:** Writing – review & editing, Project administration. **Stephanie Eagling-Peche:** Writing – review & editing, Project administration, Investigation. **Rumaya Juhari:** Writing – review & editing, Project administration. **Kufre Okop:** Writing – review & editing, Project administration. **Jennel Reyes:** Writing – review & editing, Project administration. **Viorel Babii:** Writing – review & editing, Project administration. **Lucie Cluver:** Writing – review & editing, Project administration. **Liane Peña Alampay:** Writing – review & editing, Project administration. **Marija Raleva:** Writing – review & editing, Project administration. **Frances Gardner:** Writing – review & editing, Project administration.

Role of the funding source

The ePLH pilot study is funded by the UKRI GCRF/Newton Fund, the LEGO Foundation, the philanthropic donors to the University of Oxford's COVID-19 Research Response Fund, the Oak Foundation, The Human Safety Net, UNICEF Philippines and UNICEF Malaysia and the UKRI GCRF Accelerating Achievement for Africa's Adolescents (Accelerate) Hub.

Declaration of competing interest

None.

Acknowledgement

The research team would like to acknowledge the families who took part in the ePLH study as well as the local teams from North Macedonia, Malaysia, Moldova, Montenegro, Philippines, and South Africa without whom this study would not have been possible.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2025.119524.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014. Standards for Educational and Psychological Testing. American Educational Research Association. https://doi.org/10.2307/j.ctvr43hg2.

Antony, M.M., Bieling, P.J., Cox, B., Enns, M., Swinson, R., 1998. Psychometric properties of the 42-item and 21-item versions of the depression anxiety stress scales in clinical groups and a community sample. Psychol. Assess. 10 (2), 176–181, 1040-3590/984 VOO.

Boateng, G.O., Neilands, T.B., Frongillo, E.A., Melgar-Quiñonez, H.R., Young, S.L., 2018. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front. Public Health 6, 149. https://doi.org/10.3389/fpubh.2018.00149.

Bottesi, G., Ghisi, M., Altoè, G., Conforti, E., Melli, G., Sica, C., 2015. The Italian version of the depression anxiety stress Scales-21: factor structure and psychometric properties on community and clinical samples. Compr. Psychiatry 60, 170–181. https://doi.org/10.1016/j.comppsych.2015.04.005.

Brown, G.T.L., Harris, L.R., O'Quin, C., Lane, K.E., 2017. Using multi-group confirmatory factor analysis to evaluate cross-cultural research: identifying and understanding non-invariance. Int. J. Res. Method Educ. 40 (1), 66–90. https://doi.org/10.1080/1743727X.2015.1070823.

Cai, H., Jin, Y., Liu, R., Zhang, Q., Su, Z., Ungvari, G.S., Tang, Y.-L., Ng, C.H., Li, X.-H., Xiang, Y.-T., 2023. Global prevalence of depression in older adults: a systematic review and meta-analysis of epidemiological surveys. Asian J. Psychiatr. 80, 103417. https://doi.org/10.1016/j.ajp.2022.103417.

Cianchetti, C., Pittau, A., Carta, V., Campus, G., Littarru, R., Ledda, G., Zuddas, A., Fancello, G.S., 2013. Child and adolescent behavior inventory (CABI): a new instrument for epidemiological studies and pre-clinical evaluation. Clin. Pract. Epidemiol. Ment. Health. 9, 51–61.

Cianchetti, C., Pasculli, M., Pittau, A., Campus, M.G., Carta, V., Littarru, R., Fancello, G. S., Zuddas, A., Ledda, M.G., 2017. Child and adolescent behavior inventory (CABI): standardization for age 6–17 years and first clinical application. Clin. Pract. Epidemiol. Ment. Health. 13 (1), 20–26. https://doi.org/10.2174/1745017901713010020.

Cianchetti, C., Faedda, N., Pasculli, M., Ledda, M.G., Diaz, G., Peschechera, A., Craig, F., Morelli, F., Balottin, U., Guidetti, V., Zuddas, A., Margari, L., 2020. Predictive validity for the clinical diagnosis of a new parent questionnaire, the CABI, compared with CBCL. Clin. Child Psychol. Psychiatry 25 (2), 507–519. https://doi.org/10.1177/1359104519895056.

Cluver, L., Meinck, F., Yakubovich, A., Doubt, J., Redfern, A., Ward, C., Salah, N., De Stone, S., Petersen, T., Mpimpilashe, P., Romero, R.H., Ncobo, L., Lachman, J., Tsoanyane, S., Shenderovich, Y., Loening, H., Byrne, J., Sherr, L., Kaplan, L., Gardner, F., 2016. Reducing child abuse amongst adolescents in low- and middle-income countries: a pre-post trial in South Africa. BMC Public Health 16 (1), 567. https://doi.org/10.1186/s12889-016-3262-z.

Costa, D.S., Cianchetti, C., Dewey, D., Alvim-Soares, A.M.A.-S., Kestelman, I., Da Silva, A. G., Malloy-Diniz, L.F., Miranda, D.M., De Paula, J.J., 2023. Cross-cultural adaptation, validity, and reliability of the Child and Adolescent Behavior Inventory (CABI) for use in Brazil. J. Pediatr. (Rio J) 99 (4), 413–422. https://doi.org/10.1016/j.jped.2023.02.008.

Dimitrov, D.M., 2010. Testing for factorial invariance in the context of construct validation. Meas. Eval. Couns. Dev. 43 (2), 121–149. https://doi.org/10.1177/0748175610373459.

Flora, D.B., Curran, P.J., 2004. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol. Methods 9 (4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466.

Gugerty, M.K., Karlan, D., 2018. The CART principles in more detail. In: Karlan, D., Gugerty, M.K. (Eds.), The Goldilocks challenge: right-fit evidence for the social sector. Oxford Academic, pp. 49–65. https://doi.org/10.1093/oso/9780199366088.003.0004.

Henry, J.D., Crawford, J.R., 2005. The short-form version of the depression anxiety stress scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br. J. Clin. Psychol. 44, 227–239.

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. Multidiscip. J. 6 (1), 1–55. https://doi.org/10.1080/10705519909540118.

James, S.L., Abate, D., Abate, K.H., Abay, S.M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R.S., Abebe, Z., Abera, S.F., Abil, O.Z., Abraha, H.N., Abu-Raddad, L.J., Abu-Rmeileh, N. M.E., Accrombessi, M.M.K., Murray, C.J.L., 2018. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. Lancet 392 (10159), 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7.

Jiang, L., Yan, Y., Jin, Z.-S., Hu, M.-L., Wang, L., Song, Y., Li, N.-N., Su, J., Wu, D.-X., Xiao, T., 2020. The Depression Anxiety Stress Scale-21 in Chinese hospital workers: reliability, latent structure, and measurement invariance across genders. Front. Psychol. 11, 247. https://doi.org/10.3389/fpsyg.2020.00247.

Kaku, S.M., Sibeoni, J., Basheer, S., Chang, J.P.-C., Dahanayake, D.M.A., Irarrazaval, M., Lachman, J.M., Mapayi, B.M., Mejia, A., Orri, M., Jui-Goh, T., Uddin, M.S., Vallance, I., 2022. Global child and adolescent mental health perspectives: bringing change locally, while thinking globally. Child Adolesc. Psychiatry Ment. Health 16 (1), 82. https://doi.org/10.1186/s13034-022-00512-8.

Lachman, J.M., Cluver, L., Ward, C.L., Hutchings, J., Mlotshwa, S., Wessels, I., Gardner, F., 2017. Randomized controlled trial of a parenting program to reduce the risk of child maltreatment in South Africa. Child Abuse Negl. 72, 338–351. https://doi.org/10.1016/j.chiabu.2017.08.014.

Limenih, G., MacDougall, A., Wedlake, M., Nouvet, E., 2024. Depression and global mental health in the global south: a critical analysis of policy and discourse. Int. J. Soc. Determ. Health Health Serv. 54 (2), 95–107. https://doi.org/10.1177/27551938231220230.

Little, T.D., 2023. Longitudinal Structural Equation Modeling. The Guilford Press.

MacCallum, R.C., Browne, M.W., Sugawara, H.M., 1996. Power analysis and determination of sample size for covariance structure modeling. Psychol. Methods (2).

Moussa, M.T., Lovibond, P., Laube, R., Megahead, H.A., 2017. Psychometric properties of an Arabic Version of the Depression Anxiety Stress Scales (DASS). Res. Soc. Work. Pract. 27 (3), 375–386. https://doi.org/10.1177/1049731516662916.

Murphy, H., Prandstetter, K., Ward, C.L., Hutchings, J., Kunovski, I., Tăut, D., Foran, H. M., 2024. Factor structure of the depression, anxiety and stress scale among caregivers in southeastern Europe. Fam. Relat. 73 (2), 905–920. https://doi.org/10.1111/fare.12882.

Ng, F., Trauer, T., Dodd, S., Callaly, T., Campbell, S., Berk, M., 2007. The validity of the 21-item version of the Depression Anxiety Stress Scales as a routine clinical outcome measure. Acta Neuropsychiatrica 19 (5), 304–310. https://doi.org/10.1111/j.1601-5215.2007.00217.x.

Perneger, T.V., Leplège, A., Etter, J.-F., 1999. Cross-cultural adaptation of a psychometric instrument: two methods compared. J. Clin. Epidemiol. 52 (11), 1037–1046.

Polanczyk, G.V., Salum, G.A., Sugaya, L.S., Caye, A., Rohde, L.A., 2015. Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J. Child Psychol. Psychiatry 56 (3), 345–365. https://doi.org/10.1111/jcpp.12381.

Putnick, D.L., Bornstein, M.H., 2016. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev. Rev. 41, 71–90. https://doi.org/10.1016/j.dr.2016.06.004.

R project, 2023. The R project for statistical computing. https://www.r-project.org/.

Racine, N., McArthur, B.A., Cooke, J.E., Eirich, R., Zhu, J., Madigan, S., 2021. Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: a meta-analysis. JAMA Pediatr. 175 (11), 1142. https://doi.org/10.1001/jamapediatrics.2021.2482.

Rossel, Y., 2012. Lavaan: an R package for structural equation modeling. J. Stat. Softw. 48 (2), 1–36. https://doi.org/10.18637/jss.v048.i02.

Rstudio, 2022. Rstudio. https://www.rstudio.com/products/rstudio/download/.

Schermelleh-Engel, K., Moosbrugger, H., Müller, H., 2003. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures, 8 (2), 53.

Singh, G., Prabhuappa, K.P., Eqbal, S., Singh, A.R., 2013. Depression, anxiety and stress scale: reliability and validity of Hindi adaptation. Int. J. Educ. 3 (4), 446–449.

Szabó, M., 2010. The short version of the depression anxiety stress scales (DASS-21): factor structure in a young adolescent sample. J. Adolesc. 33 (1), 1–8. https://doi.org/10.1016/j.adolescence.2009.05.014.

Ward, C.L., Wessels, I.M., Lachman, J.M., Hutchings, J., Cluver, L.D., Kassanjee, R., Nhapi, R., Little, F., Gardner, F., 2020. Parenting for Lifelong Health for Young Children: a randomized controlled trial of a parenting program in South Africa to prevent harsh parenting and child conduct problems. J. Child Psychol. Psychiatry 61 (4), 503–512. https://doi.org/10.1111/jcpp.13129.

Wolf, E.J., Harrington, K.M., Clark, S.L., Miller, M.W., 2013. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. Educ. Psychol. Meas. 73 (6), 913–934. https://doi.org/10.1177/0013164413495237.