ORIGINAL ARTICLE

Confirmatory Factor Analysis of the Kessler-6 Psychological Distress (K6) Scale in a Community Sample of People Living with Severe and Persistent Mental Illness: a Bifactor Model

Stephen R. Carter[1] · Jack C. Collins[1] · Jie Hu[2] · Claire L. O'Reilly[1] ·
Amanda J. Wheeler[2,3,4] · Sara S. McMillan[2,3,5] · Sarira El-Den[1]

Abstract

The factorial structure of the Kessler Psychological Distress Scale 10-item (K10) and 6-item (K6) has been explored, with conflicting findings across studies. Bifactor modelling has been applied to the K10, but not to the K6, which is commonly used in clinical and research settings; hence, exploring its dimensionality can demonstrate the utility of its subscales. The aim was to determine, using confirmatory factor analyses (CFA), whether the bifactor model of K6 is a suitable representation of psychological distress among a community-dwelling sample of adults living with severe and persistent mental illness. Randomised controlled trial participants ($n = 335$) completed the K6 at baseline (median score = 9/24). CFA of the various models were conducted with diagonally weighted least squares, mean and variance adjusted. The unidimensional model was rejected. The fit statistics for the correlated model and the higher-order model were identical and demonstrated comparative fit index and Tucker-Lewis index within range; however, the RMSEA values were outside the target. In the correlated model, the correlation between the factors was very high. The bifactor model had the best fit for the data. Inspection of the reliability indices from the bifactor model showed that the way participants responded to the K6 items was far more influenced by a general psychological distress factor compared with anxiety and depression factors. Tests of measurement invariance showed that participants interpreted K6 items consistently between genders but not between younger vs older age groups. Further research across larger, diverse populations should explore the suitability of suggested instrument modifications.

Keyword Kessler-6 · Psychological distress · Bifactor model · Psychometric testing · Confirmatory factor analysis

---

Stephen R. Carter and Jack C. Collins joint first-authors.

✉ Sarira El-Den
  sarira.el-den@sydney.edu.au

Extended author information available on the last page of the article

Background

Psychological distress is a state of being which is characterised by the presence of an assortment of non-specific psychological and physical signs and symptoms (Dohrenwend et al., 1980). High levels of psychological distress may be indicative of poor mental health and are often associated with mental illnesses including anxiety and depressive disorders (Dohrenwend et al., 1980; Viertiö et al., 2021). A number of scales have been developed to measure psychological distress, and such scales have been implemented in both point-of-care and widespread epidemiological surveys (Kessler et al., 2002; Viertiö et al., 2021).

The Kessler Psychological Distress Scale is widely used to measure psychological distress among the general population, as well as among targeted patient populations. There are two validated forms of the Kessler Psychological Distress Scale, namely the 10-item K10 and the six-item K6 (Kessler et al., 2002). The K10 and K6 have been extensively evaluated demonstrating their sound psychometric properties across a broad range of populations (Bougie et al., 2016; Kessler et al., 2002; Sampasa-Kanyinga et al., 2018; Sunderland et al., 2012; Umucu et al., 2021). The use of the Kessler scale is routinely recommended for screening for the risk of non-specific psychological distress in primary care. The K6 is also recommended for use as a secondary outcome measure in clinical studies, given its ability to "sensitively measure the severity of non-specific distress in the range likely to be found in clinical samples" (Kessler et al., 2002). The application of the six-item K6 is representative of a recent trend whereby mental health researchers have tried to develop shorter versions of validated measurement instruments in an attempt to reduce the time needed to complete questionnaires and minimise survey burden (El-Den et al., 2018). Despite the brevity of the K10 which is estimated to take less than 3 min to complete, the K6 is even briefer and is estimated to take less than 2 min to complete (Kessler et al., 2002).

Dimensionality of the Kessler Psychological Distress Scales

In the original validation of K10 and K6 scales, psychometric evidence demonstrated, using exploratory factor analyses and item response theory (IRT) models, that both scales measured a unidimensional construct referred to as non-specific psychological distress (Kessler et al., 2002). Since then, the K10 and K6 tools have undergone extensive psychometric testing using confirmatory factor analyses (CFA). In two-dimensional factor analytic models of K10 and K6, items probing anxiety and depression are modelled on two separate factors. Therefore, the question has arisen as to whether these tools are measuring general psychological distress which is distinct from feelings of depression and anxiety. For example, a two-dimensional CFA model of K10 demonstrated better model fit compared to the unidimensional model (Sunderland et al., 2012). With regard to K6 which is the subject of this study, fit statistics generated with the two-dimensional CFA models are superior to unidimensional models in a variety of study populations (Bessaha, 2015; Easton et al., 2017; Ko & Harrington, 2016). A better fit of the two-dimensional model lends support to the notion that K10 measures depression and anxiety as discrete constructs, and thus, the question of dimensionality remains important.

Defining the dimensionality of psychological distress may aid in the application of measurement instruments when screening for and managing psychological distress. In the context of the Kessler Psychological Distress Scales, exploring dimensionality can demonstrate whether they are useful, not only as "overall, total screeners" (Lace et al., 2020), but

potentially also for exploring the effect of the anxiety and depression subscales, in terms of identifying each illness, specifically.

One approach for exploring whether this measure of psychological distress allows for the existence of separate anxiety and depression factors is to specify a higher-order CFA model. In a higher-order model, a psychological distress factor is specified to "cause" or influence anxiety and depression factors. In a study by Peixoto et al. (2021), such a conceptualisation of the K10 had a slightly improved fit over the two-factor approach. However, also in the study by Peixoto et al. (2021), another specification of K10, namely the bifactor model, included a psychological distress factor and yielded a superior fit over other models.

Bifactor Models

Confirmatory factor analysis with bifactor modelling is an approach that is increasingly being used to test the dimensionality of multidimensional constructs in psychological research (Chen & Zhang, 2018). Comparative tests of competing models which include the bifactor model can assist researchers in making decisions regarding dimensionality, particularly for instruments used in the clinical assessment of mental health. For example, compared with other models, the bifactor model of the Hospital Anxiety and Depression Scale (HADS) demonstrates superior fit among community samples (Iani et al., 2014), as well as among specific populations such as patients living with fibromyalgia (Luciano et al., 2014). Model fit is important; however, reliability indices, which are calculated from factor loadings, further assist researchers in interpreting the degree of multidimensionality (Rodriguez et al., 2016). Iani et al. (2014) extended the HADS investigation further by reporting the reliability indices of the general factor, as well as the anxiety and depression factors arising from a bifactor model. The reliability indices of the general factor were superior to the reliability indices of the anxiety and depression factors when the effect of the general factor was partialled out. This leads to the conclusion that HADS provides a superior measure of overall negative affect than a measure of specific anxiety and depression dimensions (Iani et al. (2014).

Recent comparative tests of competing models of the K10 demonstrate the suitability and superiority of the bifactor model over other models among Australian youth (Smout, 2020). Furthermore, Smout (2020) also reported that the bifactor model demonstrated invariance of the measurement model as a function of gender and age groups. Measurement invariance implies that dimensionality remains constant between groups. In this regard, the K10 instrument has the potential to be used in future studies that seek to compare levels of general psychological distress with variations in these demographic characteristics.

The bifactor model has not been applied to the K6; therefore, the suitability of modelling psychological distress as a general factor with K6 has not been previously explored. Therefore, the dimensionality of the K6 requires further exploration to determine whether it is appropriate to score psychological distress and/or anxiety and depression separately. Furthermore, it is yet to be known whether the dimensionality of the K6 remains constant for different populations with different demographic characteristics.

The Current Study

Hence, the aim of the current study was to examine the factorial structure of the K6, using CFA, including the bifactor model, among a sample of community-dwelling adults living with severe and persistent mental illness (SPMI). Specifically, the objectives of this study were to:

- explore the dimensionality of the K6;
- determine the extent to which the factorial structure of the most superior model is consistent across groups with respect to gender, age and presence of self-reported moderate/severe anxiety, moderate/severe depression and multiple mental illnesses.

Methods

Participant Recruitment

People living with SPMI and taking an antipsychotic or mood stabiliser medication on a regular basis were recruited through community pharmacies participating in the Bridging the Gap Between Physical and Mental Illness (*PharMIbridge*) Randomised Controlled Trial (RCT) between September 2020 and May 2021 (Wheeler et al., 2020). The *PharMIbridge* RCT involved 51 randomised community pharmacies across Northern Sydney and Hunter New England in New South Wales, the Australian Capital Territory, and Regional Victoria, Australia. Further details relating to the *PharMIbridge* RCT are reported in the published protocol (Wheeler et al., 2020).

Measurement Instrument and Data Collection

The K6 asks respondents to rate how they have been feeling over the past 30 days in response to six items using a 5-point Likert scale: "none of the time" (0 points), "a little of the time" (1 point), "some of the time" (2 points), "most of the time" (3 points) and "all of the time" (4 points). Scores are then summed to give a total score out of 24, with higher scores indicating higher levels of psychological distress (Kessler et al., 2002). There is no universally accepted "cut-off" score for determining clinical levels of psychological distress or the likely presence of a mental illness. Previous studies have identified various cut-off points in different populations (Andersen et al., 2011; Cornelius et al., 2013; Kang et al., 2015; Sakurai et al., 2011). In an Australian context where this study was conducted, the Australian Bureau of Statistics states a value of 19 or above out of 30 (approximately 14–15 in the 24-point scoring system used in the current study) to be indicative of probable mental illness (Australian Bureau of Statistics, 2012). Upon their enrolment in the *PharMIbridge* RCT, consumer participants completed a baseline survey comprised of multiple validated scales, including the K6, as well as questions allowing for the collection of demographic data such as self-reported mental illness diagnoses. Cases with missing or incomplete responses to the baseline survey, at the time of extracting the data for the current study, were excluded. Age and gender were captured separately from pharmacy clinical practice software.

Data Analysis

Confirmatory Factor Analyses

CFA of the various models were conducted with diagonally weighted least squares mean and variance adjusted (WLSMV) to manage the skewness and kurtosis which is evident in ordinal data with 5 categories of responses (Foldnes & Gronneberg, 2021; Li, 2016). Model fit was evaluated using the fit indices recommended by Schreiber (2008). The

absolute fit was evaluated using the c^2, root mean error of approximation (RMSEA) and standardised root mean square residual (SRMR) which indicate how well the model fits the observed data (Hooper et al., 2008). The target was for a non-significant c^2, for RMSEA to be < 0.06 and for SRMR to be < 0.08 (Schreiber, 2008, 2017), but since higher values can be expected for RMSEA for models with low complexity (low degrees of freedom), higher values may be tolerated (Kenny et al., 2015). Relative fit indices, including the comparative fit index (CFI) and Tucker-Lewis index (TLI), were also reported (Hooper et al., 2008). Target cut-off values were ≥ 0.95 for both CFI and TLI.

The present study explores the dimensionality of the K6 by specifying alternative models and comparing these with the original model. One potentially useful model for the K6 is a reflective unidimensional model, with all items loading onto a single factor. An alternative model is a correlated model with 4 items loading onto the depression subscale and 2 items loading onto the anxiety subscale. It should be noted that the 2-item anxiety subscale requires that the loading of the two items be constrained to be equal because of the need for model identification. A higher-order model in which each of the two dimensions was specified to be indicators of a higher-order latent construct was also specified, and each of the loadings on the depression and anxiety dimensions was specified to be equal. A bifactor model was specified with an additional "general" factor, which influenced each item along with depression and anxiety dimensions. The dimensions were specified orthogonally (not correlated).

Reliability of Bifactor Scales

By apportioning the common variance of the general factor with dimensions, it is possible to determine the extent to which participants' responses to the K6 are multidimensional. For that purpose, the following reliability statistics were calculated: explained common variance (ECV) of the "general" factor; the percentage of uncontaminated correlations (PUC) representing the percentage of covariance terms which only reflect variance from the "general" factor. When ECV is high (> 0.70) and PUC is high (> 0.70), "the relative bias is slight and the common variance can be regarded as essentially unidimensional" (Rodriguez et al., 2016). If the thresholds for PUC and/ or ECV are not reached, some degree of multidimensionality exists and the degree of multidimensionality may be informed by internal consistency calculations of the "general" and other factors. Omega is used for the "general" factor, and all items are used to calculate this index of internal consistency. OmegaS is used for specific factors, and only items loading on that factor are considered. Omega Hierarchical (OmegaH) represents the percentage of systematic variance in raw total scores that can be attributed to the individual differences on the "general" factor, and evidence for unidimensionality is provided if OmegaH is high (> 0.80) (Reise et al., 2013a, 2013b). OmegaHS is calculated after partitioning out variability attributed to the "general" factor. Reise et al., (2013a, 2013b) (p. 22) suggest that when PUC is lower than the thresholds, but when ECV is reasonably high (> 0.60) and OmegaH is high (> 0.70), "the presence of some multidimensionality is not severe enough to disqualify the interpretation of the instrument as primarily unidimensional." It is also important to consider potential bias in parameter estimates arising from the use of a bifactor model compared with a unidimensional model, and for this purpose, average relative parameter bias (ARPB) was calculated. ARPB is the difference between an item's loading on the unidimensional

solution and its loading on the general factor (in the bifactor model), divided by its loading on the general factor. When ARPB is less than 10–15%, bias is acceptable and poses no concern to the interpretation of the bifactor model (Muthén et al., 1987).

Measurement Invariance

Tests of measurement invariance were performed on the best-fitting model of the K6. To determine and report the extent to which the factorial structure was consistent between groups, the guidance of Putnick and Bornstein (2016) was followed. It is recommended that, as a minimum, between-group comparisons reveal configural invariance. Configural invariance indicates that the items load onto the latent constructs in the same pattern, regardless of group. This is shown if the configural model is an overall good fit for the data (Putnick & Bornstein, 2016). For this purpose, the guidance established for assessing overall fit using the statistical tests above was used. Putnick and Bornstein (2016) suggest that if configural invariance is established, tests of metric invariance may be performed. Metric invariance indicates that each item contributes to the latent construct to a similar degree across groups. That is, the loadings are constrained to be equal across groups. However, when conducting tests of invariance on ordinal data using WLSMV, tests of metric invariance are not available when items load onto more than one factor, as occurs in a bifactor model.

The next level of invariance is scalar invariance. Scalar invariance means that mean differences in the latent constructs capture all mean differences in the shared variance of the items. Scalar invariance is demonstrated with a non-significant χ^2 difference test and when $\Delta\text{CFI} \leq 0.01$, $\Delta\text{RMSEA} \leq 0.015$ and the $\Delta\text{SRMR} \leq 0.015$ (Putnick & Bornstein, 2016). The categories of variables to be tested in the current study included the age of participants, which was divided into two age groups at the median. The other categories of variables that were to be tested included gender (male or female) and self-reported diagnosis (i.e. moderate/severe anxiety vs no moderate/severe anxiety, moderate/severe depression vs no moderate/severe depression, and one vs multiple mental illnesses).

While the presence of other self-reported mental illnesses was captured, these will not be reported in this study and tests of measurement invariance were only conducted with variation in the number of self-reported mental illnesses and the presence of self-reported depression or anxiety. The presence of self-reported anxiety and depression was selected in order to determine whether measurement invariance was present among samples with variation in these specific conditions. Furthermore, it is not recommended to conduct measurement invariance on very unbalanced group sizes, as would have been the case if modelling had been conducted on the presence of mental illnesses with small sample sizes (Yoon & Lai, 2018).

Results

Demographic Characteristics

A total of 335 consenting participants completed the K6. As can be seen in Table 1, participants had a mean age of 47.9 years (± 13.0 years) and the majority were male

(52.8%). Participants most commonly self-reported experiencing a moderate/severe depressive disorder (56.1%), followed by a moderate/severe anxiety disorder (44.5%) (Table 1). It was common for participants to self-report experiencing multiple illnesses, whereby the majority of participants (60.3%) self-reported experiencing more than one mental illness diagnosis, which may have included depression, anxiety and/or other mental illnesses (e.g. schizophrenia, bipolar disorder). The details relating to other self-reported mental illness diagnoses are beyond the scope of the current manuscript and will be reported elsewhere.

Responses to the K6

The median [interquartile range (IQR)] of the total K6 score was 9 [9], ranging from 0 to 24. Table 2 presents a descriptive analysis of participants' responses to the K6, whereby the majority of participants answered "a little of the time" for item k1 pertaining to sadness, "some of the time" for item k2 pertaining to nervousness, "some of the time" for item k3 pertaining to restlessness, "none of the time" for item k4 pertaining to hopelessness, "some of the time" for item k5 pertaining to effort and "none of the time" for item k6 pertaining to worthlessness.

Confirmatory Factor Analyses

The suitability to perform CFA was deduced from the absence of multicollinearity in the data from both samples. Analysis of the item correlation matrix showed no relationships > 0.90; therefore, multicollinearity was viewed as not problematic (Schreiber,

Table 1 Demographic characteristics ($n = 335$)

Characteristics	n	%
Gender		
Female	150	44.8
Male	177	52.8
Other or not specified	8	2.4
Age, years		
Mean (SD)	47.9 (13.0)	
Range	18.0–81.7	
< 45	136	40.6
≥ 45	199	59.4
Number of self-reported mental illnesses		
One	133	39.7
Two or more[*]	202	60.3
Self-reported mental illnesses**		
Moderate/severe depression	188	56.1
Moderate/severe anxiety disorder	149	44.5

[*]Not restricted to moderate/severe depression or anxiety disorder

**Do not add up to 100%, as multiple responses allowed

Table 2 Descriptive analysis of responses to the K6 ($n = 335$)

Item code	Item wording: *During the past 30 days, often did you feel…*	Response n (%)		Median score out of 4 [IQR]
k1	So sad nothing could cheer you up	None of the time (0):	88 (26.3%)	1 [2]
		A little of the time (1):	96 (28.7%)	
		Some of the time (2):	91 (27.2%)	
		Most of the time (3):	54 (16.1%)	
		All of the time (4):	6 (1.8%)	
k2	Nervous	None of the time (0):	48 (14.3%)	2 [2]
		A little of the time (1):	90 (26.9%)	
		Some of the time (2):	108 (32.2%)	
		Most of the time (3):	67 (20.0%)	
		All of the time (4):	22 (6.6%)	
k3	Restless or fidgety	None of the time (0):	63 (18.8%)	2 [2]
		A little of the time (1):	76 (22.7%)	
		Some of the time (2):	106 (31.6%)	
		Most of the time (3):	63 (18.8%)	
		All of the time (4):	27 (8.1%)	
k4	Hopeless	None of the time (0):	112 (33.4%)	1 [2]
		A little of the time (1):	65 (19.4%)	
		Some of the time (2):	79 (23.6%)	
		Most of the time (3):	55 (16.4%)	
		All of the time (4):	24 (7.2%)	
k5	That everything was an effort	None of the time (0):	59 (17.6%)	2 [2]
		A little of the time (1):	72 (21.5%)	
		Some of the time (2):	89 (26.6%)	
		Most of the time (3):	72 (21.5%)	
		All of the time (4):	43 (12.8%)	
k6	Worthless	None of the time (0):	115 (34.3%)	1 [2]
		A little of the time (1):	82 (24.5%)	
		Some of the time (2):	60 (17.9%)	
		Most of the time (3):	53 (15.8%)	
		All of the time (4):	25 (7.5%)	

2008). The goodness-of-fit indices for each of the four specified models are presented in Table 3.

While some of the fit statistics of the unidimensional model (Fig. 1) were within the target levels, the absolute fit $c^2 = 179$, df = 9 was significant ($p < 0.001$) and the RMSEA was very high = 0.238; hence, this model was rejected.

As expected, the fit statistics for the correlated factor model (Fig. 2) and the higher-order model (Fig. 3) were practically identical.

Compared with the unidimensional model, the two factor (correlated and the higher order) models had a much-improved fit. The absolute fit for these models $c^2 = 28.0$, df = 9 were, however, significant ($p < 0.001$). The comparative fit indices (CFI and TLI) were within range, but the RMSEA (0.079) was outside the target. It is not uncommon

Table 3 Goodness-of-fit indices for K6 using different specifications ($n=335$)

Specification	Chi-square	df	p	CFI	TLI	RMSEA	SRMR
Single factor	179	9	<0.001	0.971	0.952	0.238	0.045
Two factor	28.0	9	<0.001	0.997	0.995	0.079	0.019
Higher order	28.0	9	<0.001	0.997	0.995	0.079	0.019
Bifactor	4.89	4	0.29	1.000	0.999	0.026	0.006
Target			>0.05	>0.95	>0.95	<0.06	<0.08
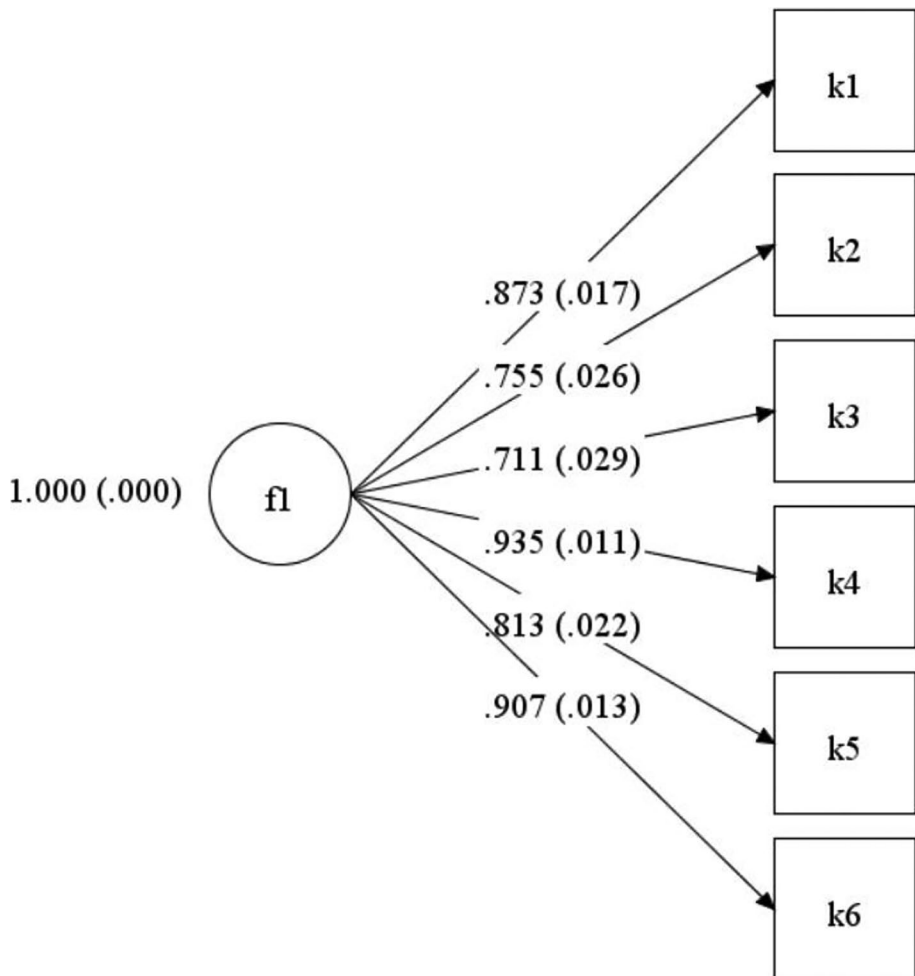
Fig. 1 Unidimensional model of the K6 ($n=335$)

for the c^2 to be significant and RMSEA to be high in small models from samples > 300, so the correlated and higher-order models should not necessarily be ignored. In the
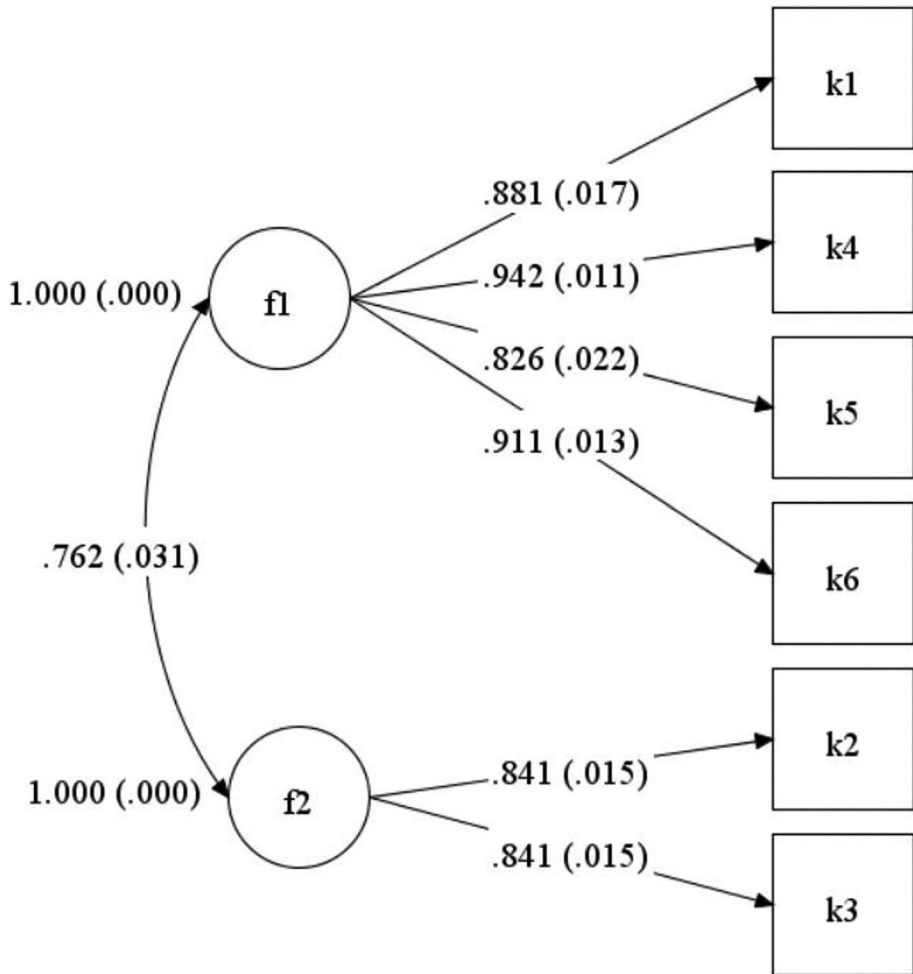
Fig. 2 Correlated factor model of the K6 ($n=335$)

correlated factor model, the correlation between the factors was positive and very high (0.762). In order to specify the higher-order model, the unstandardized factor loadings on the depression and anxiety factors were fixed to unity for reasons of identification. For the higher-order model, standardised loadings on the depression and anxiety factors were positive and very high, being 0.852 and 0.893, respectively. The much-improved fit of the correlated (and/or higher order) model over the unidimensional model supports the hypothesis that the K6 has 2 dimensions. The very high associations between the depression and anxiety factors and the theoretical understanding of the relationship between these dimensions support the investigation of a bifactor model. The bifactor model (Fig. 4) clearly had the best fit for the data, with a non-significant c^2 and all fit statistics within the target range.
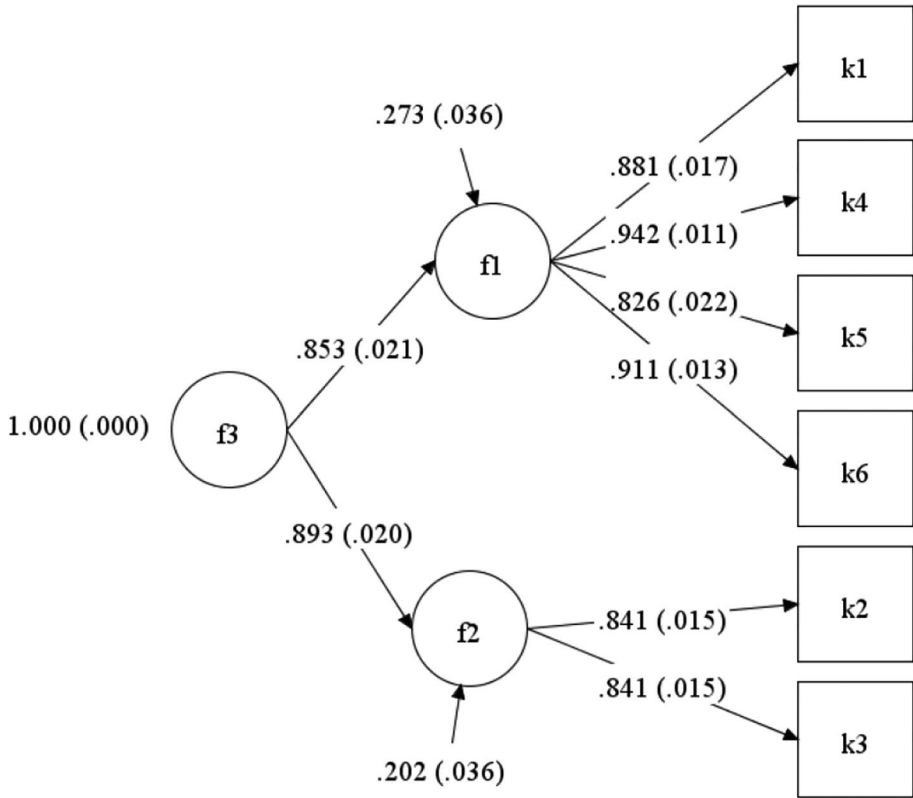
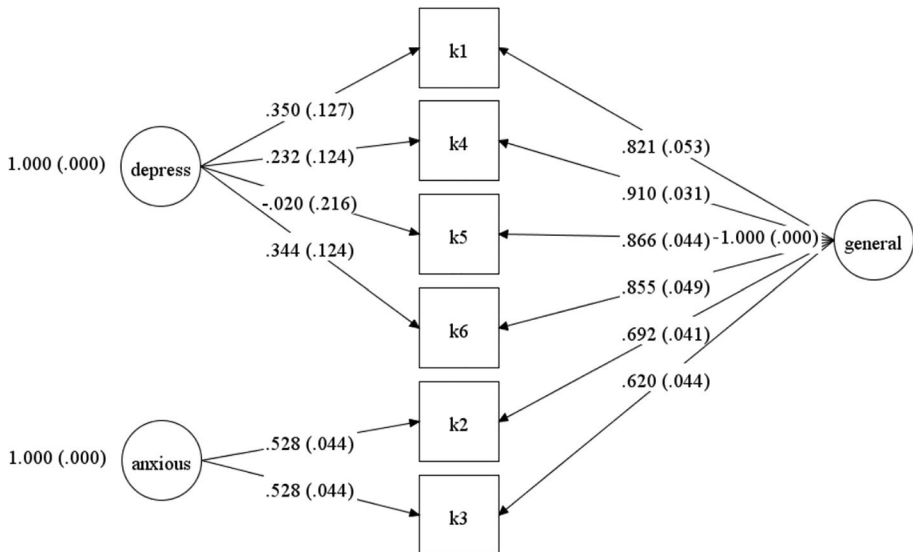**Fig. 3** Higher-order model of the K6 ($n = 335$)

Fig. 4 Bifactor model of the K6 ($n = 335$)

Reliability of Bifactor Scales

The summary statistics for the reliability of the bifactor scales are presented in Table 4. The ECV value of the general factor was high (0.819), and the overall bias in parameter estimates was low and acceptable (ARPB = 7.5%). However, the PUC value (0.533) did not exceed the threshold of > 0.7 for "essentially" unidimensionality (Rodriguez et al., 2016). There was some evidence for unidimensionality since the OmegaH value of the general factor scale (0.954) was very high, while the OmegaHS values for the depression and anxiety subscales were very low (0.061 and 0.326, respectively).

Measurement Invariance

Since the bifactor model was accepted as the most appropriate model, measurement invariance tests were conducted using this model, and the results are presented in Table 5. Each of the categories had a similar sample size, whereby $n > 100$ for all groups. The test of invariance for gender demonstrated configural invariance as all fit indices were within the target range. However, while the test of scalar invariance was accepted since the χ^2 difference test was non-significant, it should be noted that the ΔRMSEA = 0.038 and ΔSRMR = 0.017 were not within the range. The test of invariance for age (younger vs older) did not demonstrate configural invariance (RMSEA = 0.085). Exploration of the reasons why the factor structure was non-invariant between groups was conducted by performing the CFA for younger and older samples separately. The model fit was good for the younger group, $c^2 = 3.97$, df = 4, $p = 0.41$, CFI = 1.000 and RMSEA = 0.000. The model fit was not satisfactory for the older group, $c^2 = 12.81$, df = 4, $p = 0.012$, CFI = 0.996 and RMSEA = 0.123, and modification indices suggested a possible cross-loading between k2 and k6. All other tests of invariance between the self-reported presence of moderate/severe

Table 4 Summary reliability statistics of the bifactor model for the K6 ($n = 335$)

Statistic	Abbreviation	Statistical test result
Percentage uncontaminated correlations	PUC	0.533
Explained common variance (general factor)	ECV	0.819
Average relative parameter bias	ARPB	7.5%
Internal consistency		
General factor	Omega[1]	0.954
	OmegaH[2]	0.875
Anxiety factor	OmegaS[1]	0.946
	OmegaHS[2]	0.061
Depression factor	OmegaS[1]	0.831
	OmegaHS[2]	0.326

[1]Omega is used for the general factor, and all items are used to calculate this index of internal consistency. OmegaS is used for specific factors, and only items loading on that factor are considered

[2]OmegaHierarchical (OmegaH) reflects the proportion of variance in raw scores that can be attributed to the individual differences in the general factor. OmegaHS reflects the proportion of variance of a subscale after partitioning out variance attributed to the general factor

Table 5 Results of the tests of measurement invariance

Groups	Model	χ^2 (df)	P	CFI	RMSEA (90% CI)	SRMR	$\Delta\chi^2$ (Δdf)	ΔCFI	ΔRMSEA	ΔSRMR	Decision
Targets			>.05[1]	≤0.95[1]	≤0.06[1] ≤0.05[2]	≤0.08[1]		≤.01[2]	≤0.015[2]	≤0.015[2]	
Gender: male (150) vs female (177)	M1: configural invariance	4.2 (8)	0.84	1.000	.000 (.000–.050)	0.005	–	–	–	–	Accept
	M2: scalar invariance	38.5 (31)	0.166	0.999	.038 (.000–.074)	0.021	34.3 (23)	0.001	0.038	0.017	Accept (marginal)
Age: <45 (136) vs ≥45 (199)	M1: configural invariance	17.7 (8)	0.023	.998	.085 (.030–.139)	0.011	–	–	–	–	Reject
Self-reported moderate/severe depression: no (147) vs yes (188)	M1: configural invariance	13.39 (8)	0.099	0.999	.063 (.000–.121)	0.010	–	–	–	–	Accept
	M2: scalar invariance	34.49 (31)	0.304	0.999	.026 (.000–.065)	0.021	21.1 (23)	0.000	-0.037	0.011	Accept
Self-reported moderate/severe anxiety disorder: no (186) vs yes (149)	M1: configural invariance	7.743 (8)	0.459	1.000	.000 (.000–.089)	0.008	–	–	–	–	Accept
	M2: scalar invariance	23.8 (31)	0.818	1.000	.000 (.000–.037)	0.017	16.1 (23)	0.000	0.000	0.009	Accept
Number of self-reported mental illnesses: one (133) vs two or more (202)	M1: configural invariance	11.58 (8)	0.171	.999	.052 (.000–.112)	0.010	–	–	–	–	Accept
	M2: scalar invariance	31.87 (31)	0.423	1.000	.000 (.000–.037)	0.017	16.1 (23)	-0.001	-0.052	0.007	Accept

depression or not, moderate/severe anxiety or not and multiple mental illnesses or not demonstrated both configural and scalar invariance.

Discussion

This study reports the psychometric testing of the K6 in a population of community-dwelling adults living with SPMI. Previous CFA studies with the K6 have demonstrated support for both a unidimensional model and multidimensional models. Multidimensional models of the K6 have included anxiety and depression sub-dimensions. These models have been specified as a correlated model, and a hierarchical model in which a higher-order dimension of overall psychological distress "causes" these sub-dimensions. To the authors' knowledge, this is the first study to report the results of bifactor analyses on the K6. Within the sample tested in this study, the comparative fit statistics and the reliability statistics indicate that the bifactor model of the K6 is the superior representation of psychological distress. The implication is that when interpreting the K6 items, participants are predominantly influenced by overall psychological distress but also somewhat influenced by more specific anxiety and depression sub-dimensions. These interpretations are consistent for both males and females. These results are broadly consistent with recent analyses of the K10, where bifactor models produced the best fit compared with other models (Peixoto et al., 2021; Smout, 2020).

Overall, while the modelling provides evidence for a degree of multidimensionality, the multidimensionality is not severe enough to disqualify the interpretation of the instrument as primarily unidimensional (Reise et al., 2013a, 2013b). This has important implications for practice and research. The predominance of the general factor in a bifactor model means that when used in practice, the calculation of a summative score of the K6 items, as per current practice, is likely to be a reasonable estimate of an individual's level of psychological distress. In structural equation modelling, the estimation of a K6 factor score, calculated from loadings on the general factor, would provide sufficient information for estimating overall psychological distress and would not lead to significant bias. In designing analytical plans for estimating the relationship between psychological distress and other variables, it is therefore recommended that a bifactor model of the K6 be created and then causative paths (or correlations) be drawn between the general factor and other variables. Representation of the K6 in this way allows for a predominance of effect due to general distress, such as in a unidimensional model, but minimises error that may have arisen through not accounting for the coexistence of the unique effects of the depression and anxiety factors. If a unidimensional model of the K6 had been applied to the data from the present study, for example, in combination with other causative variables, it is very likely that all models would have been rejected due to the high level of error inherent within a unidimensional representation of the K6.

The presence of the depression and anxiety factors is needed to minimise error, but also to represent the way psychological distress appears to influence how participants respond to a measurement instrument. In the past, when the K6 has been modelled, an error has been minimised through allowing for the correlation of depression and anxiety factors or linking the factors through a causative hierarchical model. The hierarchical models imply that overall psychological distress "causes" depression and anxiety factors which in turn "cause" the way an individual responds to an item. Regardless, the representation of depression and anxiety subscales in such models would have meant

that it was appropriate to calculate separate factor scores for anxiety and depression factors and treat them separately in subsequent analyses. An example of the challenges arising from such an approach is evident in a study by Ko and Harrington (2016) which explored the relationship between psychological stress using K6 and suicide planning and attempts, whereby the correlated model fit the data better than the unidimensional model (no bifactor model was explored). Yet, when specified this way in a structural equation model, the authors reported that the paths between the individual anxiety and depression factors and the two behaviours lacked concurrent validity. That is, anxiety and depression treated separately did not predict behaviour in an expected manner. However, despite the fact that a unidimensional model of the K6 had significantly more error than the correlated model, in a structural equation model, psychological distress had predictable effects on both behaviours, namely, suicide planning and attempts (Ko & Harrington, 2016). Key to the appropriateness of attempting a bifactor model in that study was the observation of a high correlation between the two factors (0.848) in the correlated model. As noted by the authors, correlations (above 0.85) approached the threshold of failing tests of discriminant validity. In cases where the threshold is approached, additional tests of discriminant validity are recommended (Carter, 2016). An alternative to conducting further tests of discriminant validity when the correlations between two or more factors are very high is to explore the appropriateness of a bifactor model (Gibbons, 2014). In the current study, even though the observed correlation of depression and anxiety factors in the correlated model was only 0.762, the bifactor model was the most appropriate.

Through the application of rigorous tests of measurement invariance using the bifactor representation of the K6, it was shown that, in this cohort, males and females interpret and respond to the items in a consistent manner. This is in contrast to the results of measurement invariance testing by age, whereby the least restrictive form, configural invariance, was not evident. The bifactor model fits the data well for younger adults but not so well for older adults. Inspection of the modification indices showed that in the older cohort, items k2 ("nervous") and k6 ("worthless") share some otherwise unexplained positive co-variation (across factors). Further investigation of this possibility is recommended in future studies with different samples to determine whether this is an artefact or whether older adults interpret these items differently to younger adults. All other tests of differences between groups demonstrated scalar invariance. This suggests that the K6 could be successfully modelled and scored with a bifactor structure regardless of gender, as well as the presence of self-reported moderate/severe depression, moderate/severe anxiety or multiple mental illnesses. Nonetheless, it should be noted that this invariance may have depended on the presence of a broadly equivalent age distribution across these groups.

The specification of a bifactor model of the K6 is not without limitations. The primary issue is that in any multidimensional representation of K6, there is a lack of identification of the two-item anxiety subscale. It should be reinforced, therefore, that apart from the unidimensional model, which was rejected due to poor fit, all other models constrain items k2 and k3 to be equal. This is a very restrictive model, yet the model fit statistics show this to be a reasonable representation of dimensionality. A possible solution to this could be to create a modified K6 where one of the depression items in the present K6 is substituted with a third item loading on the anxiety subscale, potentially derived from the K10. This would generate a balance of depression and anxiety items. Alternatively, a second option could be to retain the existing K6 and add an additional item from the K10 to the anxiety subscale, creating a 7-item version of the scale. This approach avoids the need to remove items from the existing K6 and would have a limited impact on brevity. Further

sampling and analyses across diverse populations are required to explore the suitability of these potential instrument modifications.

This study is the first to apply bifactor modelling to the K6, demonstrating the suitability of this model in this population sample. However, findings should be considered in the context of potential limitations. Given the low prevalence of SPMI in the Australian adult population (in 2015, 1.1% were "estimated to be affected by a severe and persistent mental illness that requires ongoing services" (Whiteford et al., 2016)), the sample size in this study was relatively large. Nonetheless, it is important to note that the study involved a convenience sample derived from a larger RCT, and the sample size was determined by factors other than psychometric considerations, including pragmatic convenience sampling. In general, CFA requires sample sizes of > 200; however, many factors influence sample size requirements and samples as small as 60 can be appropriate for well-fitting and simple CFA models (Kyriazos, 2018). Good model fit and high factor loadings (Bessaha, 2015; Easton et al., 2017; Ko & Harrington, 2016), consistent with the current study, generally reduce sample size requirements, while the low number of items loading on a single factor (on the 2-dimensional model) may increase sample size requirements (Wolf et al., 2013). Therefore, the expected sample size of above 330 was deemed to be adequate. Sample size requirements for tests of measurement invariance were also considered suitable because group sizes were > 100 (Putnick & Bornstein, 2016). Nevertheless, further studies with a larger population and among other consumer populations are needed to determine the generalizability of findings. Finally, this study contributes to the literature by demonstrating the validity of the bifactor model of the K6 across gender and self-reported mental illness diagnoses, thereby paving the way for future research in this area.

Data Availability Data are not publicly available. Please contact the corresponding author for further information.

Declarations

Informed Consent All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all participants included in the study.

Conflict of Interest The authors declare no competing interests.

References

Andersen, L., Grimsud, A., Myer, L., Williams, D., Stein, D., & Seedat, S. (2011). The psychometric properties of the K10 and K6 scales in screening for mood and anxiety disorders in the South African stress and health study. *International Journal of Methods in Psychiatric Research, 20*(4), 215–223. https://doi.org/10.1002/mpr.351

Australian Bureau of Statistics. (2012). 4817.0.55.001 - Information paper: Use of the Kessler psychological distress scale in ABS health surveys, Australia, 2007–08 Retrieved May 20, 2022 from https://www. abs.gov.au/ausstats/abs@.nsf/lookup/4817.0.55.001chapter92007-08

Bessaha, M. L. (2015). Factor structure of the Kessler Psychological Distress Scale (K6) among emerging adults. *Research on Social Work Practice, 27*(5), 616–624. https://doi.org/10.1177/1049731515594425

Bougie, E., Arim, R., Kohen, D., & Findlay, L. (2016). Validation of the 10-item Kessler Psychological Distress Scale (K10) in the 2012 Aboriginal Peoples Survey. *Health Reports, 27*(1), 3–10.

Carter, S. R. (2016). Using confirmatory factor analysis to manage discriminant validity issues in social pharmacy research. *International Journal of Clinical Pharmacy, 38*(3), 731–737. https://doi.org/10. 1007/s11096-016-0302-9

Chen, F. F., & Zhang, Z. (2018). Bifactor models in psychometric test development. In The Wiley Handbook of Psychometric Testing (pp. 325–345). https://doi.org/10.1002/9781118489772.ch12

Cornelius, B. L. R., Groothoff, J. W., van der Klink, J. J. L., & Brouwer, S. (2013). The performance of the K10, K6 and GHQ-12 to screen for present state DSM-IV disorders among disability claimants. *BMC Public Health, 13*(1), 128–128. https://doi.org/10.1186/1471-2458-13-128

Dohrenwend, B. P., Shrout, P. E., Egri, G., & Mendelsohn, F. S. (1980). Nonspecific psychological distress and other dimensions of psychopathology. Measures for use in the general population. *Arch Gen Psychiatry, 37*(11), 1229–1236. https://doi.org/10.1001/archpsyc.1980.01780240027003

Easton, S. D., Safadi, N. S., Wang, Y., & Hasson, R. G. (2017). The Kessler psychological distress scale: Translation and validation of an Arabic version. *Health and Quality of Life Outcomes, 15*(1), 215. https://doi.org/10.1186/s12955-017-0783-9

El-Den, S., Chen, T. F., Gan, Y. L., Wong, E., & O'Reilly, C. L. (2018). The psychometric properties of depression screening tools in primary healthcare settings: A systematic review. *Journal of Affective Disorders, 225*, 503–522. https://doi.org/10.1016/j.jad.2017.08.060

Foldnes, N., & Gronneberg, S. (2021). The sensitivity of structural equation modeling with ordinal data to underlying non-normality and observed distributional forms. *Psychological Methods*. https://doi.org/ 10.1037/met0000385

Gibbons, R. (2014). Bi-factor Analysis. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 386–394). Springer, Netherlands. https://doi.org/10.1007/978-94-007-0753-5_207

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electron J Bus Res Methods, 6*, 53–60.

Iani, L., Lauriola, M., & Costantini, M. (2014). A confirmatory bifactor analysis of the hospital anxiety and depression scale in an Italian community sample. *Health and Quality of Life Outcomes, 12*(1), 84. https://doi.org/10.1186/1477-7525-12-84

Kang, Y. K., Guo, W. J., Xu, H., Chen, Y. H., Li, X. J., Tan, Z. P., Li, N., Gesang, Z. R., Wang, Y. M., Liu, C. B., Luo, Y., Feng, J., Xu, Q. J., Lee, S., & Li, T. (2015). The 6-item Kessler Psychological Distress Scale to survey serious mental illness among Chinese undergraduates: Psychometric properties and prevalence estimate. *Comprehensive Psychiatry, 63*, 105–112. https://doi.org/10.1016/j.comppsych. 2015.08.011

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods and Research, 44*(3), 486–507. https://doi.org/10.1177/ 0049124114543236

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., Walters, E. E., & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in nonspecific psychological distress. *Psychological Medicine, 32*(6), 959–976. https://doi.org/10.1017/s0033 291702006074

Ko, J., & Harrington, D. (2016). Factor structure and validity of the K6 scale for adults with suicidal ideation. *Journal of the Society for Social Work and Research, 7*(1), 43–63. https://doi.org/10.1086/685033

Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology, 9*(08), 2207.

Lace, J. W., Merz, Z. C., Grant, A. F., Emmert, N. A., Zane, K. L., & Handal, P. J. (2020). Validation of the K6 and its depression and anxiety subscales for detecting nonspecific psychological distress and need for treatment. *Current Psychology, 39*(5), 1552–1561. https://doi.org/10.1007/s12144-018-9846-2

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. https://doi.org/10. 3758/s13428-015-0619-7

Luciano, J. V., Barrada, J. R., Aguado, J., Osma, J., & García-Campayo, J. (2014). Bifactor analysis and construct validity of the HADS: A cross-sectional and longitudinal study in fibromyalgia patients. *Psychological Assessment, 26*(2), 395–406. https://doi.org/10.1037/a0035284

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52*, 431–462. https://doi.org/10.1007/BF02294365

Peixoto, E. M., Zanini, D. S., & de Andrade, J. M. (2021). Cross-cultural adaptation and psychometric properties of the Kessler Distress Scale (K10): An application of the rating scale model. *Psicologia: Reflexão e Crítica, 34*(1), 21. https://doi.org/10.1186/s41155-021-00186-9

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

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013a). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013b). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*(1), 5–26. https://doi.org/10.1177/0013164412449831

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment, 98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Sakurai, K., Nishi, A., Kondo, K., Yanagida, K., & Kawakami, N. (2011). Screening performance of K6/K10 and other screening instruments for mood and anxiety disorders in Japan. *Psychiatry and Clinical Neurosciences, 65*(5), 434–441. https://doi.org/10.1111/j.1440-1819.2011.02236.x

Sampasa-Kanyinga, H., Zamorski, M. A., & Colman, I. (2018). The psychometric properties of the 10-item Kessler Psychological Distress Scale (K10) in Canadian military personnel. *PLoS ONE, 13*(4), e0196562. https://doi.org/10.1371/journal.pone.0196562

Schreiber, J. B. (2008). Core reporting practices in structural equation modeling. *Research in Social & Administrative Pharmacy, 4*, 83–97. https://doi.org/10.1016/j.sapharm.2007.04.003

Schreiber, J. B. (2017). Update to core reporting practices in structural equation modeling. *Research in Social & Administrative Pharmacy, 13*(3), 634–643. https://doi.org/10.1016/j.sapharm.2016.06.006

Smout, M. F. (2020). The factor structure and predictive validity of the Kessler Psychological Distress Scale (K10) in children and adolescents. *Australian Psychologist, 54*(2), 102–113. https://doi.org/10.1111/ap.12376

Sunderland, M., Mahoney, A., & Andrews, G. (2012). Investigating the Factor Structure of the Kessler Psychological Distress Scale in community and clinical samples of the Australian population. *Journal of Psychopathological and Behavioral Assessment, 34*(2), 253–259. https://doi.org/10.1007/s10862-012-9276-7

Umucu, E., Fortuna, K., Jung, H., Bialunska, A., Lee, B., Mangadu, T., Storm, M., Ergun, G., Mozer, D. A., & Brooks, J. (2021). A national study to assess validity and psychometrics of the Short Kessler Psychological Distress Scale (K6). *Rehabilitation Counselling Bulletin, 65*(2), 140–149. https://doi.org/10.1177/00343552211043261

Viertiö, S., Kiviruusu, O., Piirtola, M., Kaprio, J., Korhonen, T., Marttunen, M., & Suvisaari, J. (2021). Factors contributing to psychological distress in the working population, with a special reference to gender difference. *BMC Public Health, 21*(1), 611. https://doi.org/10.1186/s12889-021-10560-y

Wheeler, A. J., O'Reilly, C. L., El-Den, S., Byrnes, J., Ware, R. S., & McMillan, S. S. (2020). Bridging the gap between physical and mental illness in community pharmacy (PharMIbridge): Protocol for an Australian cluster randomised controlled trial. *British Medical Journal Open, 10*(7), e039983–e039983. https://doi.org/10.1136/bmjopen-2020-039983

Whiteford, H., Buckingham, B., Harris, M., Diminic, S., Stockings, E., & Degenhardt, L. (2016). Estimating the number of adults with severe and persistent mental illness who have complex, multi-agency needs. *Australian and New Zealand Journal of Psychiatry, 51*(8), 799–809. https://doi.org/10.1177/0004867416683814

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. *Educational and Psychological Measurement, 73*(6), 913–934. https://doi.org/10.1177/0013164413495237

Yoon, M., & Lai, M. H. C. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling, 25*(2), 201–213. https://doi.org/10.1080/10705511.2017.1387859

Authors and Affiliations

Stephen R. Carter[1] · **Jack C. Collins**[1] · **Jie Hu**[2] · **Claire L. O'Reilly**[1] · **Amanda J. Wheeler**[2,3,4] · **Sara S. McMillan**[2,3,5] · **Sarira El-Den**[1]

1   The University of Sydney School of Pharmacy, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

2   Menzies Health Institute Queensland, Griffith University, Brisbane, Australia

3   Centre for Mental Health, Griffith University, Brisbane, Australia

4   Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand

5   School of Pharmacy and Medical Sciences, Griffith University, Gold Coast, Australia