Mental Health & Prevention 13 (2019) 100–106

Contents lists available at ScienceDirect

Mental Health & Prevention

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

Investigating the factor structure of the K10 and identifying cutoff scores denoting nonspecific psychological distress and need for treatment☆

John W. Lace[a,*], Taylor R. Greif[a], Andrew McGrath[a], Alex F. Grant[a], Zachary C. Merz[b], Carson L. Teague[a], Paul J. Handal[a]

[a] Saint Louis University, Department of Psychology
[b] University of North Carolina at Chapel Hill, Department of Physical Medicine and Rehabilitation

ARTICLE INFO

Keywords:
Mental health screening
K10
Factor analysis
Cutoff scores

ABSTRACT

While widely used in epidemiological research, studies of the Kessler Psychological Distress Scale (K10) have supported profoundly mixed results regarding its latent factor structure and optimal cutoff scores to identify clinically significant psychological distress. The present study sought to investigate the K10 and identify cutoff scores denoting clinically significant nonspecific psychological distress and need for treatment in a sample of American adults recruited online. Four hundred thirty-four (434) adults (*M* age = 37.86) participated via Amazon Mechanical Turk (MTurk) as part of a larger study and completed the International Personality Item Pool Representation of the NEO PI-R, 120-item version (IPIP-NEO-120), K10, and demographic information. Factor analyses revealed that a two-factor solution (Depression and Anxiety) fit K10 items statistically significantly better than a unidimensional model. Good convergent validity emerged between IPIP-NEO-120 Depression and Anxiety facets and K10 and its subscales (all correlation coefficients ≥ 0.63). Optimal cutoff scores denoting clinically significant nonspecific psychological distress and need for treatment were identified as: K10 (12), K10-Dep (7), and K10-Anx (3/4). The findings of the present study support the multidimensionality of the K10 in a representative sample of Americans and were the first to identify clinically validated cutoff scores for the K10's depression and anxiety subscales.

1. Introduction

Nonspecific psychological distress describes depressive, anxious, or other psychiatric symptomatology characteristic of many mental disorders (Dohrenwend, Shrout, Egri, & Mendelsohn, 1980; Drapeau, Marchand, & Beaulieu-Prévost, 2012; Mirowski & Ross, 2002) and continues to contribute to healthcare costs and disease burden worldwide (Chiu, Lebenbaum, Cheng, de Oliveira, & Kurdyak, 2017; Swartz & Jantz, 2014). Despite the point prevalence of significant psychological distress being as high as 38% (Caron et al., 2012), most individuals do not seek psychological or psychiatric treatment (Kessler et al., 2005; Mojtabai & Jorm, 2015; Mojtabai, Olfson, & Han, 2016). Thus, the development and continual validation of brief, epidemiological measures of psychological distress are paramount for psychological research and early detection of mental illness (Berwick et al., 1991; Kiely & Butterworth, 2015; Kroenke, Spitzer, & Williams, 2003). The Kessler Psychological Distress Scale (K10; Kessler et al., 2003) is a brief, 10-item screening measure of nonspecific psychological distress used widely in international epidemiological research to detect mental health dysfunction (Carrà et al., 2011; Furukawa, Kessler, Slade, & Andrews, 2003).

Reviewing current literature on the K10 reveals two primary problems. First, reported cutoff scores to detect specific disorders or otherwise operationalized distress for the K10 range notably. These cutoff scores include: ≥ 10 to detect mood or anxiety disorders in Japanese adults (Sakurai, Nishi, Kondo, Yanagida, & Kawakami, 2011); ≥ 14 to identify major depressive episodes in West African women (Baggaley et al., 2007) and French adults with alcohol use disorders (Arnaud et al., 2010); ≥ 15 to denote any mental disorder in Australian older adults (Anderson et al., 2013) and Iranian adults (Hajebi *et al.*, 2018); ≥ 16 to identify anxiety or mood disorders in South Africans (Andersen et al., 2011), Australians (Andrews & Slade, 2001), and Canadian military servicemembers (Blanc, Zamorski, Ivey, McCuaig Edge, & Hill, 2014); ≥ 19 to detect depression in Canadian older adults (Vasiliadis, Chudzinski, Gontijo-Guerra, & Préville, 2015) and Australian military servicemembers (Searle Amelia *et al.*, 2015); ≥

☆ All authors declare no conflicts of interest.

* Corresponding author: 3700 Lindell Blvd St. Louis, MO 63108.
E-mail address: lacejw@slu.edu (J.W. Lace).

https://doi.org/10.1016/j.mhp.2019.01.008
Received 14 July 2018; Received in revised form 30 November 2018; Accepted 15 January 2019
Available online 17 January 2019

20 to detect serious mental illness in New Zealanders (Oakley Browne, Wells, Scott, & McGee, 2010) and anxiety or depressive disorders in Dutch adults (Donker et al., 2010); ≥ 20/21 to denote depression in Korean older adults (Jong Won & Sun Hae, 2015); ≥ 24 to identify any mental disorder in Dutch disability applicants (Cornelius, Groothoff, van der Klink, & Brouwer, 2013) and Japanese adults (Furukawa et al., 2003); and ≥ 27 to screen for depression in Hong Kongese adolescents (Chan & Fung, 2014).

There are likely several factors contributing to the wide range of reported cutoff scores (i.e., 10 to 27). First, differential patterns of symptoms reported across cultures and clinical populations support the continued validation of the scale across appropriate samples. Second, the inconsistent operationalization of distress across studies may represent a limiting methodological factor. That is, some studies intend to use the K10 to broadly detect any mental health disorder (e.g., Anderson et al., 2013; Cornelius et al., 2013) while others seek to narrowly identify, for example, major depressive disorders (Jong Won & Sun Hae, 2015) or serious mental illness (Browne, A., Wells, Scott, & McGee, 2010). Although the K10 holds a primary epidemiological purpose, little research has attempted to detect clinically significant nonspecific psychological distress and need for treatment according to known treatment-utilization groups despite recent and remote research utilizing this methodological approach (Handal, Gist, Gilner, & Searight, 1993; Handal, Peri, & Pashak, 2015; Lace, Haeberlein, & Handal, 2019; Lace, Merz, *et al.*, 2018; Langner, 1962).

The literature is also mixed regarding the dimensionality of the K10. The scale's authors originally presented a unidimensional measure broadly capturing nonspecific psychological distress (Andrews & Slade, 2001; Kessler et al., 2003), and subsequent studies have supported a single-factor model in samples of Dutch, Turkish, and Moroccan individuals (Fassaert et al., 2009), Hong Kongese adolescents (Chan & Fung, 2014), First Nations Canadians (Bougie, Arim, Kohen, & Findlay, 2016), Palestinian social workers (Easton, Safadi, Wang, & Hasson, 2017), Danish psychiatric patients (Thelin et al., 2017), and Canadian military servicemembers (Sampasa-Kanyinga, Zamorski, & Colman, 2018). Conversely, other literature supports two primary latent factors of Depression and Anxiety for K10 items, with some arguing for further sub-delineation of these factors into dimensions of fatigue and negative affect, and nervousness and agitation, respectively (Brooks, Beard, & Steel, 2006). The two-factor structure of Depression and Anxiety has been supported in samples of community-dwelling Australians (Brooks et al., 2006), Australian psychiatric patients (Berle et al., 2010; Sunderland, Mahoney, & Andrews, 2012), Chinese adults (Bu et al., 2017) and university students (Zhou et al., 2008), Portuguese adults (Pereira et al., 2017), and Iranian psychiatric patients (Dadfar, Lester, Vahid, & Nasr Esfahani, 2016). Moreover, work in French adults with alcohol use disorders has supported a three-factor solution that captured depression, nervousness, and psychomotor agitation/anxiety (Arnaud et al., 2010).

Importantly, despite many studies supporting the multidimensionality of the K10, no study to date has attempted to identify cutoff scores against any criterion (e.g., depressive disorder, treatment-utilization) for the emerging factors (i.e., Depression and Anxiety). Recent work on the K6, the six-item short-form of the K10, in a representative sample of Americans supported a two-factor solution (of Depression and Anxiety) that fit data significantly better than a unidimensional model (Lace et al., 2018). Furthermore, cutoff scores of 4 or more on each of the two factors (K6-Depression and K6-Anxiety) adequately detected clinically significant distress and need for treatment (Lace et al., 2018).

The present study sought to replicate the methodological approaches of Lace et al. (2019) and extend it to the full-length K10 scale. It was hypothesized that a two-factor model (of depressive and anxious symptoms) would provide significantly better fit than a unidimensional model. It was also hypothesized that appropriate cutoff scores denoting clinically significant psychological distress and need for treatment, according to psychological treatment utilization status, would be identified via calculating receiver operating characteristic (ROC) curves and sensitivity, specificity, and total classification accuracy values for the K10 and its subscales.

2. Method

2.1. Participants

Participants were recruited online via the Amazon Mechanical Turk (MTurk) marketplace, shown to produce appropriately demographic representative samples (Buhrmester, Kwang, & Gosling, 2011; Casler, Bickel, & Hackett, 2013; Sprouse, 2011) and be appropriate for social scientific research (Hauser & Schwarz, 2016). Participation was restricted to those living in the United States via MTurk's internal system. The final sample included 434 individuals, with 277 (63.8%) women, 155 (35.7%) men, and 2 (0.5%) identifying as “other” gender. Ages ranged from 18 to 82, with a mean age of 37.86 (SD = 13.77). Participants' mean level of education was 15.62 years (SD = 2.66) and mean estimated annual income of $57,429 ($SD$ = $44,767). Participants were predominantly Caucasian (75%) with others identifying as Asian or Asian American (8%), Black of African American (7%), Hispanic or Latinx (6%), Multiracial (3%), or “Other” racial/ethnic identity (1%). Within the United States, most participants resided in Southeastern states (26%), with others living in the Midwest (24%), Northeast (21%), West (16%), and Southwest (13%).

2.2. Measures

K10. The K10 (Kessler et al., 2003) is a brief and efficient screener of nonspecific psychological distress used widely in epidemiological research (Carrà et al., 2011; Furukawa et al., 2003; Furukawa *et al.*, 2008). Its ten items assess frequency of common symptoms (e.g., low mood, fatigue, restlessness, nervousness) in the past 30 days. Items are scored from 0 (None of the time) to 4 (All of the time). The K10 has good psychometric properties in various samples (e.g., Anderson et al., 2013; Kessler et al., 2003; Sakurai et al., 2011). Its internal consistency was excellent in the present sample (α = 0.95).

2.2.1. IPIP-NEO-120 Anxiety and Depression facets

The IPIP-NEO-120 (Johnson, 2014) is a 120-item, public domain measure of personality. It was created to be an abbreviated version of the 300-item IPIP-NEO (Goldberg, 1999). Excellent internal consistency and construct validity have been reported for the IPIP-NEO-120 and its trait and facet scales (Johnson, 2014; Maples, Guan, Carter, & Miller, 2014), with the IPIP-NEO-120 Anxiety and Depression facets, which contain four items each, correlating at r = 0.76 with their analog NEO-PI-R scales (Johnson, 2014) and demonstrated good internal consistencies in the present sample (αs = 0.85 and 0.88, respectively). The IPIP-NEO-120 was included as a measure of convergent validity.

2.2.2. Psychological services utilization

Participants completed an item that asked, “Regarding counseling/psychological services, which best describes you?” Respondents indicated either: “I have never received counseling, therapy, and/or psychiatric/psychological services,” (n = 222; 51.2%); “I have previously received counseling, therapy, and/or psychiatric/psychological services, but no longer do,” (n = 167; 38.5%); or “I currently receive counseling, therapy, and/or psychiatric/psychological services,” (n = 45; 10.4%). Those who endorsed current service utilization were operationalized as positive for distress and need for treatment, and those who denied service utilizations were operationalized as negative for distress and need for treatment.

2.2.3. Demographics

Items assessing participants' age, race/ethnicity, years of education,

Table 1
Exploratory factor analyses on polychoric correlation matrix, and proportion of sample reporting each score.

	EFA 1*	EFA 2*		% Sample Reporting Each Score†				
	Distress.	Dep.	Anx.	Never	A Little	Sometimes	Most of the time	All of the time
Tired out for no reason	.69	.44	-	27.9	25.8	22.4	16.6	7.4
Nervous	.85	-	.67	25.6	30.9	26.0	13.4	4.1
So nervous nothing can calm you down	.91	-	.68	58.3	16.1	14.3	7.6	3.7
Hopeless	.92	.79	-	48.4	19.8	16.1	12.4	3.2
Restless/fidgety	.82	-	.92	35.7	30.0	21.2	9.9	3.2
So restless cannot sit still	.80	-	.87	53.2	22.1	17.1	5.5	2.1
Depressed	.88	.91	-	38.5	25.6	19.1	11.8	5.1
Everything is an effort	.82	.85	-	35.7	25.8	17.1	15.7	5.8
So sad cannot cheer up	.92	.88	-	56.7	18.9	13.8	8.1	2.5
Worthless	.92	.91	-	56.9	16.8	13.6	8.1	4.6
% Variance	75.5	75.5	7.3	-	-	-	-	-
Alpha	.95	.93	.89	-	-	-	-	-

Note.
* $n = 206$. Factor loadings below $|.40|$ were suppressed.
† $N = 434$.

geographic region of residence, approximate household income, and other demographic variables were included.

2.3. Procedures

Approval was obtained from the affiliated university's institutional review board (IRB #29,412). Participants were recruited and participated in the survey online via MTurk. Participants reviewed an IRB-approved outline and rationale for the study and voluntary decision to participate in and complete the survey served as informed consent. Several data validity checks were embedded throughout the protocol, and only those with valid responses to all validity items were retained for analysis. Eighty individuals responded in an invalid manner and were excluded from analyses. No missing data were identified.

The protocol was completed in approximately 15 min per participant. Financial compensation of $0.30 was provided upon completion via electronic transfer through MTurk. The amount of $0.30 was deemed appropriate as it was greater than compensation provided in other research of similar length (Difallah, Catasta, Demartini, Ipeirotis, & Cudr-Maroux, 2015).

2.4. Statistical analyses

The present study sought to replicate the methodology and extend the findings of recent work on the K6 (Lace et al., 2018). Of note, K10 items are ordinal in nature, and appropriate nonparametric methods were preferred. Median test revealed that K10 scores were not significantly different between men and women, $\chi^2 = 0.34$ $p = .63$; thus, all data were collapsed. The combined sample was randomly separated into two subsamples to properly conduct exploratory and confirmatory factor analyses (EFAs & CFAs; Izquierdo, Olea, & Abad, 2014).

Traditional factor analytic techniques relying on Pearson product-moment correlations and multivariate normality (e.g., maximum likelihood) were inappropriate due to the ordinal nature of K10 items. Unweighted least squares (ULS) was chosen as the estimation method for all factor analyses (i.e., exploratory *and* confirmatory) due to its appropriateness for these type of data (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Izquierdo et al., 2014; Jin, 2012; Koğar & Koğar, 2015) and recent use in similar research (Lace et al., 2018).

In the first subsample ($n_{EFA} = 206$), polychoric correlation matrices were calculated using FACTOR (Lorenzo-Seva & Ferrando, 2006; Lorenzo-Seva & Ferrando, 2017) and a series of EFAs using ULS were performed to extract two theoretically plausible models identified in reviewed literature. The first was a forced-one-factor model (e.g., Kessler et al., 2003) and the second was a forced-two-factor model (e.g., Bu et al., 2017) with promax (oblique) rotation. Factor loadings ≥ 0.40 were considered significant.

In the second subsample ($n_{CFA} = 228$), CFAs using ULS were conducted in Amos. Available fit indices were interpreted against conservative cutoff criteria. These were: $\chi^2/df \leq 3$ (Schreiber *et al.*, 2006); Goodness of Fit Index (GFI) ≥ 0.95 (Shevlin & Miles, 1998); Normed Fit Index (NFI) ≥ 0.95; and Standardized Root Mean Square Residual (SRMR) ≤ 0.08 (Hu & Bentler, 1999; Tabachnik & Fidell, 2013). χ^2-difference test was performed to determine which model demonstrated significantly better fit.

In the combined sample ($N = 434$), convergent validity was investigated through Spearman's rho correlation coefficients between the K10 and its subscales and IPIP-NEO-120 facets measuring anxiety and depression. A multivariate analysis of variance (MANOVA) was conducted with psychological services utilization groups as the independent variable and K10 and IPIP-NEO-120 facets as the dependent variables.

The combined sample was used to determine appropriate cutoff scores for the K10 and its identified subscales. Of 434 participants, 222 denied ever using psychological services and were operationalized as negative for distress (TX-). Forty-five (45) reported current psychological service utilization and were operationalized as positive for distress and need for treatment (TX+). The remaining 167 participants who noted previous but not current psychological service utilization were excluded from only these analyses. Thus, 267 participants were included to determine cutoff scores. ROC curves were calculated and areas under the curve (AUCs) were compared according to Hanley and McNeil (1983). Sensitivity, specificity, Youden's *J* (Youden, 1950), and total classification accuracy were calculated for various cutoff scores.

3. Results

3.1. Exploratory factor analyses

Table 1 displays exploratory factor analysis results. Good sampling adequacy was found for the polychoric correlation matrix of K10 items (Kaiser-Meyer-Olkin, KMO = 0.93) and Bartlett's test of sphericity ($\chi 2 = 1835.80$, $p < .001$) suggested items may be efficiently analyzed. The forced-one-factor EFA (Table 1, EFA 1) accounted for 75.5% of the variance, with item loadings ranging from 0.69 to 0.92.

The forced-two-factor model (Table 1, EFA 2) accounted for 82.7% of the variance. No items cross-loaded. Factors correlated strongly at 0.79. Six items loaded onto the first factor ($\alpha = 0.93$) and appeared to capture symptoms of depression (e.g., hopelessness, low mood, fatigue). Four items loaded onto the second factor ($\alpha = 0.89$) and appeared to measure symptoms of anxiety (e.g., restlessness, nervousness).

Table 2
Fit indices from confirmatory factor analyses.

	χ^2	*df*	χ^2/df	GFI	NFI	SRMR
One-Factor Model	66.31*	20	3.32	.99†	.99†	.06†
Two-Factor Model^	50.05*	21	2.38†	.99†	.99†	.05†

Note. $n = 228$.
* $p < .01$.
† Met strict predetermined criterion denoting excellent fit. ^Model with significantly better fit. GFI = Goodness of Fit Index. NFI = Normed Fit Index. SRMR = Standardized Root Mean Square Residual.

3.2. Confirmatory factor analyses

Table 2 displays fit indices for confirmatory factor analyses. The unidimensional model for K10 items revealed good fit. The significant χ^2 value (66.31, $p < .01$) may be reflective of large sample size and did not outright indicate poor fit. Although three fit indices surpassed predetermined cutoff criteria (GFI, NFI, and SRMR), its χ^2/df failed to meet predetermined cutoff criterion.

The two-factor model for K10 items was in accordance with factor loadings from EFA 2 (Table 1), with latent factors allowed to correlate. Notably, the significant χ^2 value (50.05, $p < .01$) may also be related to large sample size. The two-factor model revealed good-to-excellent fit, with *all* fit indices surpassing conservative predetermined cutoff criteria. Additionally, a χ^2-difference test (Schermelleh-Engel, Moosbrugger, & Müller, 2003; Werner & Schermelleh-Engel, 2010) between the two tested models was significant ($\Delta\chi^2 = 16.26$, $p < .001$), indicating that the two-factor model provided statistically significantly better fit than the unidimensional model.

3.3. Convergent validity

All Spearman's rho correlation coefficients were statistically significant. Results revealed that NEO-120-Anx correlated strongly with K10 Total (ρ = 0.67), K10-Dep (ρ = 0.63), and K10-Anx (ρ = 0.65). Similarly, NEO-120-Dep correlated strongly with K10 Total (ρ = 0.76), K10-Dep (ρ = 0.79), and K10-Anx (ρ = 0.63).

The MANOVA was statistically significant, $F(8, 856) = 8.47$, $p < .001$, Wilks' $\Lambda = 0.86$, partial eta-squared = 0.07. As can be seen in Table 3, those currently receiving psychological services reported significantly higher K10, K10-Dep, and IPIP-NEO-120 facet scores than those previously receiving services and never receiving services. While those currently receiving services reported significantly greater K10-Anx scores than those never receiving treatment, they did not report significantly higher K10-Anx scores ($M = 5.87$) than those previously receiving services ($M = 4.91$, $p = .19$), although visual inspection of mean differences showed a trend toward significance. Additionally, those previously receiving services reported significantly higher K10, K10-Dep, K10-Anx, and IPIP-NEO-120 facet scores than those never receiving services.

Table 4
Sensitivity, specificity, and overall classification accuracy for cutoffs based on total K10 score.

Cutoff Score	Sensitivity	Specificity	Youden's *J*	Accuracy
8	.84	.59	.43	.63
9	.82	.66	.48	.69
10	.82	.69	.51	.71
11	.82	.70	.52	.72
12	**.80**	**.72**	**.52**	**.73**
13	.76	.72	.48	.73
14	.64	.74	.38	.72
15	.60	.76	.36	.73

Note. $n = 267$ for these analyses. The row describing the cutoff score of 12 is displayed in boldface as it represented the ideal cutoff score in the present sample. Youden's J = (sensitivity + specificity – 1) (Youden, 1950).

3.4. Selection of cutoff scores

Results from ROC analyses revealed moderate-to-large (Douglas, Guy, Reeves, & Weir, 2005) AUCs for the K10 total (AUC = 0.78, $p < .001$), K10-Dep (AUC = 0.79, $p < .001$), and K10-Anx (AUC = 0.71, $p < .001$), suggesting that each variable was significantly better than chance at classifying participants as TX+ or TX-. Comparison of AUC values (Hanley & McNeil, 1983) revealed that both the K10 total and K10-Dep were significantly greater than the K10-Anx ($ps < 0.05$). AUC values for K10 total and K10-Dep were not significantly different ($p > .05$). Various cutoff scores are displayed in Table 4 and 5. For the total K10, the optimal cutoff score appeared to be 12, yielding sensitivity of 0.80, specificity of 0.72, and overall classification accuracy of 0.73. For K10-Dep, the optimal cutoff score appeared to be 7, with sensitivity of 0.82, specificity of 0.70, and total classification accuracy of 0.72. For K10-Anx, a cutoff score of 3 yielded good sensitivity (0.80), low specificity (0.56) and low classification accuracy (0.60), and a cutoff score of 4 provided sensitivity of 0.71, specificity of 0.63, and total classification accuracy of 0.64. While a score of 3 yielded greater Youden's J, a score of 4 provided better classification accuracy. Notably, despite no significant differences among AUC values, the K10 total and K10-Dep appeared to relatively better identify clinically significant distress than the K10-Anx.

4. Discussion

The present study strove to contribute to the literature by providing a psychometric investigation of the K10 in a representative sample of Americans that replicated and extended recent work on the K6 (Lace et al., 2018). The findings of the present study provide clarity on research regarding the K10's latent dimensionality. Results of exploratory and confirmatory factor analyses supported a two-factor solution comprised of Depression and Anxiety, which is in line with other work on the K10 in various populations (e.g., Bu et al., 2017; Zhou et al., 2008). This finding is in contrast with research supporting the unidimensionality of the K10 across studies (e.g., Andrews & Slade,

Table 3
Group differences among K10 and IPIP-NEO-120 variables.

Scale	*M* (*SD*) Never ($n = 222$)	Previously ($n = 167$)	Current ($n = 45$)	Sig. Differences*
K10 Total	8.38 (8.53)	12.41 (10.02)	17.40 (8.35)	N < P < C
K10-Dep	5.09 (5.43)	7.50 (6.50)	11.53 (5.70)	N < P < C
K10-Anx	3.29 (3.44)	4.91 (4.12)	5.87 (3.61)	N < P, C
IPIP-NEO-120-Dep	8.54 (3.80)	10.22 (4.22)	12.62 (3.87)	N < P < C
IPIP-NEO-120-Anx	10.61 (3.98)	12.60 (4.09)	14.22 (4.01)	N < P < C

Note. $N = 434$.
* $p < .05$. Never/N = Never used psychological services. Previously/P = Previously used psychological services, but no longer do. Current/C = Currently use psychological services.

Table 5
Sensitivity, specificity, and overall classification accuracy for cutoffs for K10-Dep and K10-Anx.

Cutoff Score	Sensitivity	Specificity	Youden's *J*	Accuracy
K10-Dep				
5	.89	.59	.48	.64
6	.82	.65	.47	.68
7	**.82**	**.70**	**.52**	**.72**
8	.76	.71	.47	.72
9	.69	.74	.43	.73
K10-Anx				
2	.87	.42	.29	.50
3	**.80**	**.56**	**.36**	**.60**
4	**.71**	**.63**	**.34**	**.64**
5	.64	.68	.32	.68
6	.53	.76	.29	.72

Note. The rows describing the ideal cutoff scores of 7 for K10-Dep and 3/4 for K10-Anx are displayed in boldface. Youden's J = (sensitivity + specificity – 1) (Youden, 1950).

2001; Kessler et al., 2003; Thelin et al., 2017). While a unidimensional solution demonstrated good fit according to confirmatory fit indices alone, results revealed that the two-factor solution provided *statistically significantly* better and improved fit. Further, excellent construct validity was identified through correlational analyses with trait-like anxious and depressive tendencies (ρs ranged from 0.63 to 0.79), such that the two identified subscales likely assess underlying constructs theorized. Additionally, psychological service utilization groups tended to show significant differences on every dependent variable, with those currently receiving psychological services reporting significantly higher scores on K10, K10-Dep, K10-Anx, and IPIP-NEO-120 facets than those never receiving services. These findings further supported the construct validity of the K10 and its subscales. Thus, it is likely appropriate to interpret the K10 as measuring depressive and anxious symptoms related to nonspecific, transdiagnostic psychological distress. Methodologically, these factor analytic findings support the routine comparison of competing factor structures instead of relying on fit indices in isolation for determining appropriate fit.

Furthermore, the present study was the first to identify clinically validated cutoff scores denoting psychological distress and need for treatment for the K10's subscales (K10-Dep and K10-Anx), despite extensive literature supporting its two-factor structure. Results provided preliminary support for the K10-Dep and K10-Anx at detecting nonspecific psychological distress and need for treatment that may provide clinical utility above and beyond the K10 total score. Notably, the K10-Anx scale at a cutoff score of 3 yielded good sensitivity (0.80) and lower specificity (0.56) and classification accuracy (0.60), and a cutoff score of 4 yielded relatively lower sensitivity (0.71) and marginally improved specificity (0.63) and classification accuracy (0.64). At both a score of 3 and 4, K10-Anx appeared less sensitive and accurate than did K10-Dep or K10 total, and statistical comparisons of AUC values corroborated this finding. It may be that K10 items are better suited to detect depressive symptoms than anxious symptoms, despite both being considered dimensions of nonspecific psychological distress. Future research should seek to validate the K10-Anx at detecting both nonspecific psychological distress and screening those with diagnosed anxiety disorders.

The K10 and its subscales may offer clinical and research utility in several ways. Related to clinical utility, the K10 may function as an efficient screener for significant psychological distress in various healthcare settings. Given its brevity, good psychometric properties, and freely available use, the K10 may be well-suited for use across clinical situations, including in primary or integrated care facilities, in order to alert clinicians to patients who may benefit from a more comprehensive, in-depth mental health assessment or referral for psychological or psychiatric care. That is, clinics may consider adopting policies and procedures wherein all patients complete a K10 before meeting with their provider. The clinician may then be able to provide immediate feedback regarding the likelihood of clinically significant distress related to anxious and depressive symptoms, discuss the patient's concerns in detail in the office, and offer appropriate clinical referrals and treatment options. In this way, the K10 may play a role in detecting psychological distress early and increasing the likelihood of patients receiving appropriate care in primary and integrated care settings.

Related to research applications, the K10 may find continued use in internet-based research. That is, given its conciseness and ease of administration in an online format, it may be a preferred measure for researchers conducting epidemiological studies or for those searching for a freely available measure of psychological distress for inclusion in a larger battery of self-report assessment measures. Additionally, the K10 may be well-suited for use in long-term projects needing brief, efficient, well-validated measures of nonspecific psychological distress over time. For example, clinical researchers may utilize the K10 in tracking changes in levels of nonspecific distress over the course of an intervention or other longitudinal studies. In sum, the K10 and its subscales identified above appear to have several clinical and research applications.

The present study had two primary limitations. First, the criterion against which the K10 was validated in the present study was *current* psychological treatment utilization. This methodological consideration was chosen to remain consistent with decades of previous literature validating screening measures for psychological distress and need for treatment (e.g., Handal et al., 1993; Handal et al., 2015; Lace et al., 2019; Langner, 1962) and due to recent research using this criterion in investigation of the K6 (Lace et al., 2018). The "known groups" methodology employed was deemed appropriate for determining the epidemiological utility of the K10 given the purpose of identifying nonspecific psychological distress and need for treatment.

Moreover, the prevalence of current psychological service utilization in the present sample was 10.4%, which is in line with estimates of distress prevalence and treatment-seeking. Epidemiological research on treatment-seeking behavior in community-based samples suggests that approximately 20% to 30% of those experiencing clinically significant distress seek professional help (Kessler et al., 2005). To illustrate, if approximately one-third of individuals are suffering with psychological distress or otherwise identified mental illness (Drapeau et al., 2012; Kilkkinen et al., 2007) and only between 20% and 30% of these individuals seek treatment (Kessler et al., 2005), then between 8% and 11% of the total population may currently be receiving psychological treatment at any given time. This estimation is very similar to the identified rate of treatment-seeking behavior in the present sample. Furthermore, 37.6% of participants in the present study scored above the identified cutoff score of 12 on K10 total to denote clinically significant distress and need for treatment. Notably, this figure is also in line with previous research suggesting that approximately 38% of individuals are currently experiencing "high psychological distress" (Caron et al., 2012, p. 5). In all, these findings are consistent with epidemiological work on nonspecific psychological distress and treatment-seeking behavior in representative samples and suggest that the criterion variable of treatment-utilization was appropriate given the study's methodological aims.

A second limitation regards the use of the IPIP-NEO-120 as a measure with which the K10 and its subscales were correlated, as it was designed to assess *trait*-like tendencies toward anxious and depressive presentation rather than *state*-like symptoms. That is, the IPIP-NEO-120 was not designed to detect state-like symptoms of depression and anxiety, but rather reflective of personality tendencies toward depressive and anxious expression (Johnson, 2014). However, the IPIP-NEO-120 facets also demonstrated significant differences among psychological service utilization groups, suggesting that those who are currently receiving services reported higher level of anxious and depressive

personality tendencies than those previously or never using psychological services. Thus, the authors believed it was appropriate to include the IPIP-NEO-120 to remain methodologically consistent with previous research on the K6 (Lace et al., 2018) and due to the highly overlapping relationships between trait neuroticism and development and expression of psychopathology (Mahaffey, Watson, Clark, & Kotov, 2016).

Future research should seek to continually validate the K10, its subscales (including the K6), and proposed cutoff scores in representative samples *and* specific populations of interest. Furthermore, while not the aim of the present study, other research should investigate the diagnostic utility of the K10's subscales against psychometrically robust symptom checklists for depressive and anxious disorders (e.g., GAD-7; Spitzer, Kroenke, Williams, & Löwe, 2006; PHQ-9; Kroenke, Spitzer, & Williams, 2003), self-reported diagnostic history, and structured clinical interviews. Additionally, the K10 and its subscales may be investigated in the context of disorder severity, such that those of mild, moderate, or severe expressions of psychopathology may be detected by unique cutoff scores. The authors encourage those who have access to archival datasets in which participants were structurally clinically interviewed *and* the K10 was administered to investigate the K10's subscales. In conclusion, while the present study was the first to offer preliminary validity for the K10's subscales, future work is needed to continually investigate other applications.

Conflict of interest statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Compliance with Ethical Standards:

Funding

This study received no funding.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

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

Conflict of interest

Regarding manuscript entitled "Investigating the Factor Structure of the K10 and Identifying Cutoff Scores Denoting Nonspecific Psychological Distress and Need for Treatment," all authors have no conflicts of interest to disclose.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mhp.2019.01.008.

References

Andersen, L. S., Grimsrud, A., Myer, L., Williams, D. R., Stein, D. J., & 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.

Anderson, T. M., Sunderland, M., Andrews, G., Titov, N., Dear, B. F., & Sachdev, P. S. (2013). The 10-Item Kessler psychological distress scale (K10) as a screening instrument in older individuals. *The American Journal of Geriatric Psychiatry, 21*(7), 596–606. hTtps://doi.org/10.1016/j.jagp.2013.01.009.

Andrews, G., & Slade, T. (2001). Interpreting scores on the Kessler psychological distress scale (K10). *Australian and New Zealand Journal of Public Health, 25*(6), 494–497. https://doi.org/10.1111/j.1467-842X.2001.tb00310.x.

Arnaud, B., Malet, L., Teissedre, F., Izaute, M., Moustafa, F., Geneste, J., & Brousse, G. (2010). Validity study of Kessler's psychological distress scales conducted among patients admitted to French emergency department for alcohol consumption–related disorders. *Alcoholism: Clinical and Experimental Research, 34*(7), 1235–1245. https://doi.org/10.1111/j.1530-0277.2010.01201.x.

Baggaley, R. F., Ganaba, R., Filippi, V., Kere, M., Marshall, T., Sombie, I., & Patel, V. (2007). Detecting depression after pregnancy: The validity of the K10 and K6 in Burkina Faso. *Trop Med Int Health, 12*(10), 1225–1229. https://doi.org/10.1111/j.1365-3156.2007.01906.x.

Berle, D., Starcevic, V., Milicevic, D., Moses, K., Hannan, A., Sammut, P., & Brakoulias, V. (2010). The factor structure of the Kessler-10 questionnaire in a treatment-seeking sample. *The Journal of Nervous and Mental Disease, 198*(9), 660–664. https://doi.org/10.1097/NMD.0b013e3181ef1f16.

Berwick, D. M., Murphy, J. M., Goldman, P. A., Ware, J. E., Jr., Barsky, A. J., & Weinstein, M. C. (1991). Performance of a five-item mental health screening test. *Med Care, 29*(2), 169–176.

Blanc, S., Zamorski, M., Ivey, G., McCuaig Edge, H., & Hill, K. (2014). How much distress is too much on deployed operations? Validation of the Kessler psychological distress scale (K10) for application in military operational settings. *Military Psychology, 26*(2), 88–100. https://doi.org/10.1037/mil0000033.

Bougie, E., Arim, R. G., Kohen, D. E., & Findlay, L. C. (2016). Validation of the 10-item Kessler psychological distress scale (K10) in the 2012 aboriginal peoples survey. *Health Rep, 27*(1), 3–10.

Brooks, R. T., Beard, J., & Steel, Z. (2006). Factor structure and interpretation of the K10. *Psychological Assessment, 18*(1), 62–70. https://doi.org/10.1037/1040-3590.18.1.62.

Bu, X.-q., You, L.-m., Li, Y., Liu, K., Zheng, J., Yan, T.-b., & Zhang, L.-f. (2017). Psychometric properties of the Kessler 10 scale in Chinese parents of children with cancer. *Cancer Nursing, 40*(4), 297–304. https://doi.org/10.1097/ncc.0000000000000392.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk:A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3–5. https://doi.org/10.1177/1745691610393980.

Caron, J., Fleury, M.-J., Perreault, M., Crocker, A., Tremblay, J., Tousignant, M., . . ., & Daniel, M. (2012). Prevalence of psychological distress and mental disorders, and use of mental health services in the epidemiological catchment area of Montreal South-West. *BMC Psychiatry, 12*(1), 183. https://doi.org/10.1186/1471-244x-12-183.

Carrà, G., Sciarini, P., Segagni-Lusignani, G., Clerici, M., Montomoli, C., & Kessler, R. C. (2011). Do they actually work across borders? Evaluation of two measures of psychological distress as screening instruments in a non Anglo-Saxon country. *European Psychiatry, 26*(2), 122–127. https://doi.org/10.1016/j.eurpsy.2010.04.008.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior, 29*(6), 2156–2160. https://doi.org/10.1016/j.chb.2013.05.009.

Chan, S. M., & Fung, T. C. T. (2014). Reliability and validity of K10 and K6 in screening depressive symptoms in Hong Kong adolescents. *Vulnerable Children and Youth Studies, 9*(1), 75–85. https://doi.org/10.1080/17450128.2013.861620.

Chiu, M., Lebenbaum, M., Cheng, J., de Oliveira, C., & Kurdyak, P. (2017). The direct healthcare costs associated with psychological distress and major depression: A population-based cohort study in Ontario, Canada. *PLOS ONE, 12*(9), e0184268. https://doi.org/10.1371/journal.pone.0184268.

Cornelius, B. L., Groothoff, J. W., van der Klink, J. J., & 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. https://doi.org/10.1186/1471-2458-13-128.

Dadfar, M., Lester, D., Vahid, A. K. M., & Nasr Esfahani, M. (2016). The psychometric parameters of the Farsi form of the Kessler psychological distress scale (K10) in psychiatric outpatients. *International Journal of Medical Research & Health Sciences, 5*(2), 1–7.

Difallah, D. E., Catasta, M., Demartini, G., Ipeirotis, P. G., & Cudr-Maroux, P. (2015). The dynamics of micro-task crowdsourcing: The case of Amazon MTurk. *Paper presented at the Proceedings of the 24th International Conference on World Wide Web*.

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. *Archives of General Psychiatry, 37*(11), 1229–1236. https://doi.org/10.1001/archpsyc.1980.01780240027003.

Donker, T., Comijs, H., Cuijpers, P., Terluin, B., Nolen, W., Zitman, F., & Penninx, B. (2010). The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. *Psychiatry Research, 176*(1), 45–50. hTtps://doi.org/10.1016/j.psychres.2009.01.012.

Douglas, K. S., Guy, L. S., Reeves, K. A., & Weir, J. (2005). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography*. 1–90 Available online: http://kdouglas.files.wordpress.com/2006/04/annotate10-24nov2008.pdf, Originally published in 2005. Current version updated through November 2008.

Drapeau, A., Marchand, A., & Beaulieu-Prévost, D. (2012). Epidemiology of psychological distress. In L. LAbate (Ed.). *Mental illnesses - understanding, prediction and control* (pp. 105–134). IntechOpen.

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.

Fassaert, T., De Wit, M. A. S., Tuinebreijer, W. C., Wouters, H., Verhoeff, A. P., Beekman, A. T. F., & Dekker, J. (2009). Psychometric properties of an interviewer-administered version of the Kessler Psychological distress scale (K10) among Dutch, Moroccan and Turkish respondents. *International Journal of Methods in Psychiatric Research, 18*(3), 159–168. https://doi.org/10.1002/mpr.288.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with

ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(4), 625–641. https://doi.org/10.1080/10705510903203573.

Furukawa, T. A., Kawakami, N., Saitoh, M., Ono, Y., Nakane, Y., Nakamura, Y., ..., & Kikkawa, T. (2008). The performance of the Japanese version of the K6 and K10 in the World Mental Health Survey Japan. *Int J Methods Psychiatr Res, 17*(3), 152–158. https://doi.org/10.1002/mpr.257.

Furukawa, T. A., Kessler, R. C., Slade, T., & Andrews, G. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychological Medicine, 33*(2), 357–362. https://doi.org/10.1017/S0033291702006700.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe, 7*(1), 7–28.

Hajebi, A., Motevalian, A., Amin-Esmaeili, M., Rahimi-Movaghar, A., Sharifi, V., Hoseini, L., ..., & Mojtabai, R. (2018). Adaptation and validation of short scales for assessment of psychological distress in Iran: The Persian K10 and K6. *Int J Methods Psychiatr Res, 27*(3), e1726. https://doi.org/10.1002/mpr.1726.

Handal, P. J., Gist, D., Gilner, F. H., & Searight, H. R. (1993). Preliminary validity for the Langner symptom survey and the brief symptom inventory as mass-screening instruments for adolescent adjustment. *Journal of Clinical Child Psychology, 22*(3), 382–386. https://doi.org/10.1207/s15374424jccp2203_9.

Handal, P. J., Peri, A., & Pashak, T. J. (2015). Calibration of the Langner symptom survey for the college population. *Current Psychology, 34*(2), 389–400. https://doi.org/10.1007/s12144-014-9264-z.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148*(3), 839–843. https://doi.org/10.1148/radiology.148.3.6878708.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods, 48*(1), 400–407. https://doi.org/10.3758/s13428-015-0578-z.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

Izquierdo, I., Olea, J., & Abad, F. J. (2014). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema, 26*(3), 395–400. https://doi.org/10.7334/psicothema2013.349.

Jin, R. (2012). *Sample size in exploratory factor analysis with ordinal data (Doctoral dissertation).* Gainesville, FL: University of Florida.

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51*(Supplement C), 78–89. https://doi.org/10.1016/j.jrp.2014.05.003.

Jong Won, M., & Sun Hae, L. (2015). Validation of the K6/K10 scales of psychological distress and their optimal cutoff scores for older Koreans. *The International Journal of Aging and Human Development, 80*(3), 264–282. https://doi.org/10.1177/0091415015590316.

Kessler, R. C., Demler, O., Frank, R. G., Olfson, M., Pincus, H. A., Walters, E. E., & Zaslavsky, A. M. (2005). Prevalence and treatment of mental disorders, 1990 to 2003. *New England Journal of Medicine, 352*(24), 2515–2523. https://doi.org/10.1056/NEJMsa043266.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., & Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Arch Gen Psychiatry, 60*(2), 184–189.

Kiely, K. M., & Butterworth, P. (2015). Validation of four measures of mental health against depression and generalized anxiety in a community based sample. *Psychiatry Research, 225*(3), 291–298. hTtps://doi.org/10.1016/j.psychres.2014.12.023.

Kilkkinen, A., Kao-Philpot, A., O'Neil, A., Philpot, B., Reddy, P., Bunker, S., & Dunbar, J. (2007). Prevalence of psychological distress, anxiety and depression in rural communities in Australia. *Australian Journal of Rural Health, 15*(2), 114–119. https://doi.org/10.1111/j.1440-1584.2007.00863.x.

Koğar, H., & Koğar, E. Y. (2015). Comparison of different estimation methods for categorical and ordinal data in confirmatory factor analysis. *Journal of Measurement and Evaluation in Education and Psychology, 6*(2), 351–364.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2003). The Patient Health Questionnaire-2: Validity of a two-item depression screener. *Med Care, 41*(11), 1284–1292. https://doi.org/10.1097/01.mlr.0000093487.78664.3c.

Lace, J. W., Haeberlein, K. A., & Handal, P. J. (2019). Multidimensionality of the Langner symptom survey and replication of a cutoff score in emerging adults. *Psychological Reports, 122*(1), 340–359. https://doi.org/10.1177/0033294118755112.

Lace, J. W., Merz, Z. C., Grant, A. F., Emmert, N. A., Zane, K. L., & Handal, P. J. (2018). Validation of the K6 and its depression and anxiety subscales for detecting non-specific psychological distress and need for treatment. *Current Psychology*. https://doi.org/10.1007/s12144-018-9846-2.

Langner, T. S. (1962). A twenty-two item screening score of psychiatric symptoms indicating impairment. *Journal of Health and Human Behavior, 3*(4), 269–276. https://doi.org/10.2307/2948599.

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods, 38*(1), 88–91. https://doi.org/10.3758/BF03192753.

Lorenzo-Seva, U., & Ferrando, P. J. (2017). FACTOR ver. 10.8.02. Retrieved from http://psico.fcep.urv.es/utilitats/factor/index.html.

Mahaffey, B. L., Watson, D., Clark, L. A., & Kotov, R. (2016). Clinical and personality traits in emotional disorders: Evidence of a common framework. *Journal of Abnormal Psychology, 125*(6), 758–767. https://doi.org/10.1037/abn0000187.

Maples, J. L., Guan, L., Carter, N. T., & Miller, J. D. (2014). A test of the international personality item pool representation of the revised NEO personality inventory and development of a 120-item IPIP-based measure of the five-factor model. *Psychological Assessment, 26*(4), 1070–1084. https://doi.org/10.1037/pas0000004.

Mirowsky, J., & Ross, C. E. (2002). Selecting outcomes for the sociology of mental health: Issues of measurement and dimensionality. *Journal of Health and Social Behavior, 43*, 152–170.

Mojtabai, R., & Jorm, A. F. (2015). Trends in psychological distress, depressive episodes and mental health treatment-seeking in the United States: 2001–2012. *Journal of Affective Disorders, 174*, 556–561. hTtps://doi.org/10.1016/j.jad.2014.12.039.

Mojtabai, R., Olfson, M., & Han, B. (2016). National trends in the prevalence and treatment of depression in dolescents and young adults. *Pediatrics, 138*(6), 1–12. https://doi.org/10.1542/peds.2016-1878.

Oakley Browne, A. M., Wells, J. E., Scott, K. M., & McGee, M. A. (2010). The Kessler psychological distress scale in Te Rau Hinengaro: The New Zealand mental health survey. *Australian & New Zealand Journal of Psychiatry, 44*(4), 314–322. https://doi.org/10.3109/00048670903279820.

Pereira, A., Monteiro, S., Oliveira, C. A., Vagos, P., Bártolo, A., & Jardim, J. (2017). Reliability and factor structure of the 10-item Kessler psychological distress scale (K10) among Portuguese adults. Retrieved from http://hdl.handle.net/11328/2144.

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.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research, 8*(2), 23–74.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research, 99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338.

Searle Amelia, K., Van Hooff, M., McFarlane Alexander, C., Davies Christopher, E., Fairweather-Schmidt, A. K., Hodson Stephanie, E., & Steele, N. (2015). The validity of military screening for mental health problems: Diagnostic accuracy of the PCL, K10 and AUDIT scales in an entire military population. *International Journal of Methods in Psychiatric Research, 24*(1), 32–45. https://doi.org/10.1002/mpr.1460.

Shevlin, M., & Miles, J. N. V. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences, 25*(1), 85–90. https://doi.org/10.1016/S0191-8869(98)00055-5.

Spitzer, R. L., Kroenke, K., Williams, J. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The gad-7. *Archives of Internal Medicine, 166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092.

Sprouse, J. (2011). A validation of Amazon mechanical turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods, 43*(1), 155–167. https://doi.org/10.3758/s13428-010-0039-7.

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 Psychopathology and Behavioral Assessment, 34*(2), 253–259. https://doi.org/10.1007/s10862-012-9276-7.

Swartz, J. A., & Jantz, I. (2014). Association between nonspecific severe psychological distress as an indicator of serious mental illness and increasing levels of medical multimorbidity. *American Journal of Public Health, 104*(12), 2350–2358. https://doi.org/10.2105/AJPH.2014.302165.

Tabachnik, B. G., & Fidell, L. (2013). *Using Multivariate Statistics.* Boston, MA: Allyn & Bacon.

Thelin, C., Mikkelsen, B., Laier, G., Turgut, L., Henriksen, B., Olsen, L. R., & Arnfred, S. (2017). Danish translation and validation of Kessler's 10-item psychological distress scale - K10. *Nord J Psychiatry, 71*(6), 411–416. https://doi.org/10.1080/08039488.2017.1312517.

Vasiliadis, H.-M., Chudzinski, V., Gontijo-Guerra, S., & Préville, M. (2015). Screening instruments for a population of older adults: The 10-item Kessler psychological distress scale (K10) and the 7-item generalized anxiety disorder scale (GAD-7). *Psychiatry Research, 228*(1), 89–94. hTtps://doi.org/10.1016/j.psychres.2015.04.019.

Werner, C., & Schermelleh-Engel, K. (2010). Deciding between competing models: Chi-square difference tests. Retrieved from https://perma.cc/2RTR-8XPZ.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32–35 doi:10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.

Zhou, C. C., Chu, J., Wang, T., Peng, Q., He, J., Zheng, W., & Xu, L. (2008). Reliability and validity of 10-item Kessler scale (K10) Chinese version in evaluation of mental health status of Chinese population. *Chin J Clin Psychol, 16*(6), 627–629.