SA Journal of Industrial Psychology
ISSN: (Online) 2071-0763, (Print) 0258-5200

Clinical validation of brief mental health scales for use in South African occupational healthcare

Authors:
Charles H. van Wijk[1,2]
Jarred H. Martin[3]
David J.F. Maree[3]

Affiliations:
[1]Institute for Maritime Medicine, Simon's Town, South Africa

[2]Department of Global Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

[3]Department of Psychology, Faculty of Humanities, University of Pretoria, Pretoria, South Africa

Corresponding author:
Charles van Wijk,
chvanwijk@gmail.com

Dates:
Received: 01 Mar. 2020
Accepted: 15 July 2021
Published: 30 Aug. 2021

How to cite this article:
Van Wijk, G.H., Martin, J.H., & Maree, D.J.F. (2021). Clinical validation of brief mental health scales for use in South African occupational healthcare. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde, 47*(0), a1895. https://doi.org/10.4102/sajip.v47i0.1895

Orientation: South Africa carries a high burden of mental ill-health. Screening to identify individuals for further referral is emerging as one pathway to promote access to mental health interventions. Existing occupational health surveillance infrastructure may be a useful mechanism for clinical mental health screening.

Research purpose: This study explored the clinical validity of a range of brief mental health measures in the context of occupational health surveillance.

Motivation for the study: To meaningfully screen for mental health as part of occupational health surveillance, tools are required that are empirically validated, clinically useful, locally available and practical to administer.

Research approach/design and method: Workers (n = 1816), recruited through workplace occupational health surveillance programmes, completed the Patient Health Questionnaire-9, Brief Symptom Inventory 18-somatisation subscale, Generalised Anxiety Disorder scale-7, Primary Care Post-Traumatic Stress Disorder Screen, Intense (panic-like) anxiety scale and CAGE scale and partook in a diagnostic interview with a clinical psychologist.

Main findings: Basic psychometric characteristics were reported, including confirmatory factor analyses, measurement invariance, internal consistencies and socio-demographic effects. Clinical utility was explored through receiver operating/operator characteristics curve analyses, and calculations of positive and negative predictive values, as well as sensitivity and specificity. These indicators provided evidence of clinical validity in the study context.

Practical/managerial implications: The findings support the use of psychological screening as a brief, practicable and easily accessible mode of occupational mental health support.

Contribution/value-add: This article presented evidence of structural and criterion validity for these scales and described their clinical application for practical use in occupational mental health surveillance.

Keywords: CAGE; clinical screening; GAD-7; occupational health surveillance; occupational mental health; PC-PTSD-5; PHQ-9.

Introduction

Orientation

The provision of mental health services in South Africa (SA) faces serious challenges, in part because of severe resource constraints (Docrat, Besada, Cleary, Daviaud, & Lund, 2019). It is in addressing the gap between the mental health needs and the availability of providers that screening, with the aim of identifying individuals for further referral or support, has emerged as one pathway to promote access to individuals who require mental health interventions. Existing occupational health surveillance mechanisms may be a useful vehicle for clinical mental health screening. Yet, to screen meaningfully and efficiently, tools are required that are empirically validated, clinically useful, locally available and practical to administer, and in doing so, meet the ethico-legal standards for the use of psychologically orientated screening measures in South African workplaces (Employment Equity Act, 1998). The combined set of brief mental health scales, as described under Methods, has not yet been validated for use in a South African occupational health surveillance context.

Research purpose and objectives

This study set out to explore the clinical validity of a set of brief mental health measures in the specific context of occupational health surveillance initiatives. The objectives of the study were two-fold:

- Objective 1: To provide evidence of structural validity for the screening measures, with specific reference to unidimensionality and measurement invariance.
- Objective 2: To provide evidence of criterion validity, as well as clinical utility, with reference to screening efficacy in accurately identifying risk for poor mental health in the occupational healthcare context.

Literature review

The most comprehensive local survey of common mental disorders (CMD) in SA to date was the SA Stress and Health (SASH) study (Herman et al., 2009; Stein et al., 2008; Tomlinson, Grimsrud, Stein, Williams, & Myer, 2009), which, for example, estimated the 12-month prevalence of major depressive disorder (MDD) at 4.9%, any anxiety disorder at 8.1%, and substance use disorders at 5.8%, amongst other conditions. However, it has subsequently been suggested that the SASH data may have underestimated the actual prevalence of CMD in SA (Jacob & Coetzee, 2018).

Recent SA studies estimated higher prevalence in high-risk (i.e. for adverse mental health outcomes) occupational groups (Rossouw, Seedat, Emsley, Suliman, & Hagemeister, 2013; South African Police Service, 2016; Van Wijk, Cronje, & Meintjes, 2020; Ward, Lombard, & Gwebushe, 2006). A recent general workplace sample (Van Wijk, Martin, & Meintjes, 2021) reported a current time-point prevalence of 3.6% for MDD, 3.0% for Generalised Anxiety Disorder (GAD), 1.2% for post-traumatic stress disorder (PTSD), and 4.0% for alcohol use disorder (AUD).

Within SA, the provision of mental healthcare is associated with a number of challenges, predominantly around insufficient resources and disparities in access to care (Docrat et al., 2019). This includes a lack of care providers (e.g. psychiatric nurses, psychiatrists, psychologists), as well as infrastructure (e.g. psychiatric beds, capacity at primary healthcare clinics). For example, less than 5% of the SA health budget is spent on mental health and less than 8% of that at primary healthcare level. Furthermore, in the public sector, there are only 0.31 psychiatrists, 0.79 psychologists and 1.83 social workers per 100 000 population, and it is estimated that only 1 in 10 people in SA living with a mental health condition receive the care they need (Docrat et al., 2019). Budgetary and capacity limitations in the public health system often translate to an ongoing lack of mental healthcare infrastructure, including mechanisms to appropriately identify and respond to mental illness in the population at large.

In the workplace, poor mental health is associated with significant costs, both human and economic (Mall et al., 2015; Schoeman, 2017; Stander, Bergh, Miller-Janson, De Beer, & Korb, 2016; Zungu, 2013). For example, major depression and anxiety disorders are estimated to cause a loss of earnings of R54 000 per affected adult per year, with the total annual cost to the South African economy amounting to more than R40-billion annually (Schoeman, 2017). Other reports suggest that one in four employed workers or managers have been formally diagnosed with depression (Stander et al., 2016). Another local estimate implicated substance abuse in 50% of SA workplace accidents (McCann et al., 2011). In addition, international studies reported an increased risk for workplace accidents and injuries where CMD are present (Hilton & Whiteford, 2010; Kessler, Lane, Stang, & Van Brunt, 2009; Palmer, D'Angelo, Harris, Linaker, & Coggon, 2014; Soares, Gelmini, Brandão, & Silva, 2018). Apart from the economic costs, poor mental health has personal implications, from the demands on individuals to manage their conditions, to reduced personal accomplishment and sense of self-worth, as well as the challenges of dealing with perceived stigma at work.

Employers have developed greater awareness of the deleterious effects of poor mental health on human resource management and corporate success and have sought to establish mechanisms to actively address this through Employer Assistance or Employer Well-being Programmes. Parallel to these programmes, occupational health surveillance is already a statutory requirement in the workplace and continues to expand to include some form of mental health monitoring. Adapting existing occupational health surveillance infrastructure for mental health screening could become an efficient point of entry to enable the streaming of identified individuals towards appropriate mental health services, and thus substantially contribute to the identification of need and timeous referral for intervention. Mental health screening could include any process aimed at identifying, amongst groups of people, individuals at risk for poor mental health, in order to allow the streaming of those in need towards further assessment or intervention. Screening is typically brief and aimed at identifying need for referral, rather than at making a diagnosis.

Rationale and aim

To integrate the identification of mental health needs into existing occupational health surveillance infrastructure, appropriate screening tools are required. A large number of potential measures are available (cf. Mulvaney-Day et al., 2017, for review), but few have been adequately studied in local SA settings and particularly in the context of occupational healthcare provision. Neither their fair and unbiased use (Employment Equity Act, 1998) nor their clinical validity (i.e. accuracy in identifying risk), have been established in this context. There is agreement that validation is a constant process, involving a continuum of evidentiary support, including evidence of internal structures and effects of context and sample characteristics (AERA, 2014; EFPA, 2013; Schaap & Kekana, 2016). Before any screening scales can therefore be used with confidence, evidence of validity

in local settings is required. This study investigated six measures, described in the Methods section, which speak to the mental health conditions most often encountered in the workplace and aimed to provide evidence of validity for this specific set of tools in the specific context of occupational mental health surveillance.

Research design

Research approach

This study followed a cross-sectional survey design and quantitatively analysed data obtained through the completion of psychological scales.

Research method

Participants

Participants were recruited through workplace occupational health surveillance programmes and invited to complete the questionnaire booklet and partake in an interview during their annual occupational health assessments. All participants (n = 1816) gave written informed consent to the process.

Participants were included in the study if they had a minimum of 9 years of formal schooling. This was to ensure a level of English proficiency sufficient to complete the mental health measures described here. Their ages ranged from 20 to 60 (M = 33.8, ± 8.2) and 37.4% were women. All participants were in full-time salaried employment and comprised skilled and semi-skilled workers. Further composition of the sample (home language, occupational background) is provided in Table 1. The data were from workers across multiple sites in different provinces. The sample does not necessarily represent any specific community or industry in SA, as it was a convenience sample, recruited from a range of industries and geographical locations. Data collection took place from January to December 2019.

Instruments

The six measures, described here, purport to measure mental health conditions most commonly encountered in the workplace. Most of them have internationally reported evidence of validity and have previously been studied in local populations (often in primary healthcare) or are currently being used in SA industry. All were available in the public domain and could be readily reproduced. Table 2 provides an overview of the internal reliability and diagnostic accuracy of these scales, whilst Table 3 provides an overview of dimensionality and confirmatory factor analysis data. All the scales are screening tools and not intended to confirm clinical diagnoses.

TABLE 1: Sample composition in terms of home language and occupational field.

Language	%	Occupational field*	%
English	18.9	Administrative/clerical	13.6
Setswana	11.6	Hospitality/catering	5.0
Sesotho	10.7	Marine officers	11.1
Sepedi	8.0	Navy personnel	11.5
isiXhosa	12.0	Qualified technicians (mechanical/electrical)	12.4
isiZulu	11.8	Radar/sonar/remote underwater vehicle operators	9.3
Tsivenda	3.8	Technical assistants (not formally qualified)	5.6
isiTsonga	1.8	Security services	14.2
Afrikaans	17.4	Other/unknown	17.3
Ndebele	1.7	-	-
Siswati	2.0	-	-
Other	0.3	-	-

*, Occupational groups comprising < 5% of the total sample were collapsed into the 'Other' category.

Patient Health Questionnaire-9

Major depressive disorder is a syndrome characterised by severe and persistent low mood, profound sadness, sense of despair or anhedonia (APA, 2021). The Patient Health Questionnaire-9 (PHQ-9) is a screening, diagnostic and monitoring tool that measures the severity of depression in primary care settings (Kroenke, Spitzer, & Williams, 2001; Spitzer, Kroenke, & Williams, 1999). Each item is scored on a range from 0 (*not at all*) to 3 (*nearly every day*), with higher scores indicating higher levels of depression. The 9-item scale has high internal consistency (see Table 2) and good test–retest reliability in Western (r = 0.84, Kroenke et al., 2001) and African (r = 0.90, Adewuya, Ola, & Afolabi, 2006; r = 0.75, Weobong et al., 2009) samples. The scale rates the frequency of symptoms and Question 9 screens for the presence and duration of suicide ideation. It has a follow up, non-scored question that assigns weight to the degree to which depressive symptoms have affected a patient's level of functioning. This 10th item was not completed at this study's participating sites.

The United States of America validation study (Kroenke et al., 2001) and follow up United Kingdom validation study (Gilbody, Richards, & Barkham, 2007) demonstrated good sensitivity and specificity for major depression, whilst studies from sub-Saharan Africa report a range of sensitivities with good specificity (see Table 2; Adewuya et al., 2006; Bhana, Rathod, Selohilwe, Kathree, & Petersen, 2015; Cholera et al., 2014; Pence et al., 2012). Acceptable international and local Cronbach's alphas have consistently been reported (see Table 2 for summary). In low- and middle-income contexts, a score of ≥ 10 was previously recommended as a positive screen for depression (Adewuya et al., 2006; Akena et al., 2012; Manea, Gilbody, & McMillan, 2012) and a similar value was considered optimal in the occupational health setting (Volker et al., 2016). The PHQ-9 has extensively been used in international workplace studies (e.g. Asami, Goren, & Okumura, 2015; Jain et al., 2013; Newcomb et al., 2016).

Evidence of criterion validity have previously been reported from samples from sub-Saharan Africa (Adewuya et al., 2006; Botha, 2011; Cholera et al., 2014; Pence et al., 2012), together with data indicating a unidimensional structure of the scale (see Table 3; Botha, 2011; Kigozi, 2020). Whilst most studies support the unidimensionality of the PHQ-9, other models have also been suggested (cf. Lamela, Soreira, Matos, & Morais, 2020, for review). A systematic review found evidence of measurement invariance across all available studies (Lamela et al., 2020).

TABLE 2: Overview of internal reliability and diagnostic accuracy data.

Source	Sample description	*n*	Alpha	ROC		PPV (%)	NPV (%)	SENS (%)	SPEC (%)	Optimal cut-off
			α	AUC	95%CI					
Patient Health Questionanaire-9 (9 items)										
Spitzer et al. (1999)	Primary healthcare samples	3585	-	-	-	-	-	75.0	90.0	-
Kroenke et al. (2001)	Primary healthcare samples	6000	0.89	-	-	-	-	88.0	88.0	≥ 10
Adewuya et al. (2006)	Nigerian university students	512	0.85	-	-	75.0	99.6	84.6	99.4	≥ 10
Gilbody et al. (2007)	Primary healthcare samples	93		0.94	0.89–0.98	-	-	91.7	78.3	≥ 10
Weobong et al. (2009)	Ghana, post-natal sample	160	0.79	0.90	0.81–0.98	-	-	-	-	-
Botha (2011)	SA, community sample	185	0.76	-	-	-	-	-	-	-
Manea et al. (2012)	Meta-analysis (pooled data)	5782	-	-	-	-	-	85.0	89.0	≥ 10
Pence et al. (2012)	Cameroon, HIV positive sample	400	-	-	-	12.0	98.0	27.0	94.0	≥ 10
Cholera et al. (2014)	SA, primary healthcare clinics	397	-	0.88	0.83–0.92			78.7	83.4	
Bhana et al. (2015)	SA, primary healthcare clinics (Setswana version)	676	0.76	0.85	0.82–0.88	50.0	94.0	49.0	94.0	≥ 9
Volker et al. (2016)	Adult employees (occupational health setting)	170	-	0.90	0.85–0.94	-	-	86.1	78.4	≥ 10
Wilson, Wissing and Schutte (2018)	SA, community sample (Setswana version)	376	0.71	-	-	-	-	-	-	-
Maroufizadeh, Omani-Samani, Almasi-Hashiani, Amini, and Sepidarkish (2019)	People with infertility: Iran (Persian version)	539	0.85	-	-	-	-	-	-	-
Kigozi (2020)	SA, patients with Tuberculosis	208	0.84	-	-	-	-	-	-	-
Brief Symptom Inventory 18 – somatisation scale (6 items)										
Derogatis (2001)	Original American norm sample		0.74	-	-	-	-	-	-	-
Meachen et al. (2008)	Persons with traumatic brain injury (2 samples)	176	0.75	-	-	-	-	-	-	-
		81	0.61	-	-	-	-	-	-	-
Petkus et al. (2010)	Home-bound older adults	142	0.69	-	-	-	-	-	-	-
Wang et al. (2010)	Drug users	710	0.84	-	-	-	-	-	-	-
Franke et al. (2017)	General population sample (German version)	2516	0.82	-	-	-	-	-	-	-
Generalised anxiety disorder-7 (7 items)										
Spitzer et al. (2006)	Primary care sample	2739	0.92	0.91		29.0	99.0	89.0	82.0	≥ 10
Kroenke et al. (2007)	Primary care sample	965	-	0.91	0.87–0.94	-	-	89.0	82.0	≥ 10 (GAD only)
				0.86	0.84–0.89	-	-	73.0	85.0	≥ 9 (any anxiety disord)
Lowe et al. (2008)	General population sample (German version)	5036	0.89	-	-	-	-	-	-	-
García-Campayo et al. (2010)	Spanish version	212	0.94	0.96	0.93–0.99	92.9	87.6	86.8	93.4	≥ 10
Simpson et al. (2014)	Perinatal women	240	-	0.71	-	50.5	80.5	61.3	72.7	≥ 13
Barthel et al. (2014)	Pregnant women: Ivory coast	647	0.69	-	-	-	-	-	-	-
	Pregnant women: Ghana	395	0.67	-	-	-	-	-	-	-
Zhong et al. (2015)	Pregnant women: Peru (Spanish version)	946	0.89	0.75	0.68–0.80	3.0	99.4	73.0	67.0	≥ 7
Bezuidenhout (2018)	SA, working adults	644	0.92	-	-	-	-	-	-	-
Omani-Samani et al. (2018)	People with infertility: Iran (Persian version)	539	0.88	-	-	-	-	-	-	-
Henn and Morgan (2019)	SA, working adults	529	0.92	-	-	-	-	-	-	-
Primary Care PTSD screen for DSM-5 (5 items)										
Marx et al. (2014)	Veterans	140	-	-	-	81.0	78.0	83.0	75.0	≥ 4
Prins et al. (2016)	Veterans	398	-	0.94	0.91–0.97	51.0	99.0	95.0	85.0	≥ 3
Jung et al. (2018)	Korean version	252	0.87	0.90	-	68.1	94.1	90.8	77.6	≥ 3
Bovin et al. (2021)	Veterans	396	0.96	0.93	0.90–0.96	48.0	97.0	90.0	80.0	≥ 3
CAGE scale (4 items)										
Claassen (1999)	SA, rural community (Afrikaans version)	96	-	-	-	-	-	100	78.0	≥ 2
Dhalla and Kopec (2007)	Systematic review	-	-	-	-	-	-	71.0	90.0	≥ 2
Williams (2014)	Review	-	-	-	-	-	-	91.0	77.0	≥ 2
Vissoci et al. (2018)	Tanzania, clinical samples (Swahili version)	190	0.76	-	-	-	-	-	-	-

ROC, receiver operating/operator characteristic; AUC, area under the curve; 95% CI, 95% confidence interval; PPV, positive predictive value; NPV, negative predictive value; SENS, sensitivity; SPEC, specificity; SA, South Africa; HIV, human immunodeficiency virus; GAD, Generalised Anxiety Disorder; PTSD, post-traumatic stress disorder; DSM, Diagnostic and statistical manual of mental disorders (5th ed.).

Brief Symptom Inventory 18-somatisation scale

The Brief Symptom Inventory 18 (BSI-18) (Derogatis, 2001) is an 18-item self-report checklist developed as a brief screen for psychological symptoms in medical patients. Items are rated on a 5-point Likert-type scale ranging from 0 (*not at all*) to 4 (*extremely*), with psychometric properties and cultural influences well established (Asner-Self, Schreiber, & Marotta, 2006; Petkus et al., 2010; Petrowski, Schmalbach, Jagla,

TABLE 3: Overview of fit indices for determining dimensionality of study scales.

Source	Scale	Dimensionality	χ^2	*df*	χ^2/ *df* ratio	GFI	AGFI	CFI	TLI	NFI	SRMR	RMSEA	(90% CI)
Lowe et al. (2008)	GAD-7	Unidimensional	314.1**	14	-	-	-	0.990	0.990	-	-	0.065	0.059–0.072
García-Campayo et al. (2010)	GAD-7	Unidimensional	-	-	2.8	0.960	0.910	0.990	-	0.980	-	0.080	-
Zhong et al. (2015)	GAD-7	Unidimensional	-	-	-	-	-	0.969	-	-	0.046	0.051	0.043–0.059
Bezuidenhout (2018)	GAD-7	Unidimensional	49.38**	14	-	-	-	0.960	0.940	-	0.040	0.118	0.080–0.150
Omani-Samani et al. (2018)	GAD-7	Unidimensional	-	-	3.48	0.970	-	0.990	-	-	0.031	0.068	-
Botha (2011)	PHQ-9	Unidimensional	-	-	-	0.950	-	-	-	-		0.070	-
Maroufizadeh et al. (2019)	PHQ-9	Unidimensional	-	-	4.29	-	-	0.980	-	-	0.044	0.078	-
Kigozi (2020)	PHQ-9	Unidimensional	-	-	-	-	-	0.982	0.976	-	-	0.062	0.032–0.089
Vissoci et al. (2018)	CAGE	Unidimensional	3.02	2	-	-	-	0.99	0.98	-	-	0.05	-
Current study			-	-	-	-	-	-	-	-	-	-	-
PHQ-9	9 items	Unidimensional	109.235**	27	4.05	-	-	0.947	0.929	-	0.036	0.041	0.033–0.049
GAD-7	7 items	Unidimensional	61.645**	14	4.40	-	-	0.967	0.951	-	0.029	0.048	0.036–0.061
PC-PTSD-5	5 items	Unidimensional	34.046**	5	6.81	-	-	0.976	0.953	-	0.059	0.057	0.040–0.075
CAGE	4 items	Unidimensional	8.256*	2	4.13	-	-	0.998	0.995	-	0.040	0.042	0.015–0.073

χ^2, Chi square; *df*, degrees of freedom; GFI, goodness of fit index; AGFI, adjusted goodness of fit index; CFI, comparative fit index; TLI, Tucker–Lewis Index; NFI, normed fit index; SRMR, standardised root mean square residual; RMSEA, root mean square error of approximation; PHQ-9, Patient Health Questionnaire - 9; GAD-7, Generalised Anxiety Disorder scale - 7; PC-PTSD-5, Primary care Post traumatic stress disorder screen - 5; 90% CI, 90% confidence interval.

*, $p < 0.05$; **, $p < 0.001$.

Franke, & Brähler, 2018). Application studies found that the BSI-18 is suitable for measuring psychological distress and comorbidities in patients with different mental and somatic illnesses (Adams, Boscarino, & Galea, 2006; Carlson et al., 2004; Recklitis, Blackmon, & Chang, 2017). The BSI-18-S contains six items comprising the somatisation scale, with generally acceptable alphas reported (see Table 2; Franke et al., 2017; Meachen, Hanks, Millis, & Rapport, 2008; Petkus et al., 2010; Wang et al., 2010). Women report higher scores than men (Franke et al., 2017). The BSI-18-S does not directly tap any specific CMD; rather, it was included in the present study based on previous observations that (South) Africans may use somatisation as expression of psychological distress (Draguns & Tanaka-Matsumi, 2003; Swartz, 1998).

Generalised Anxiety Disorder scale -7

Generalised anxiety disorder is a syndrome characterised by excessive and uncontrollable anxiety and worry about a range of concerns (APA, 2021). The Generalised Anxiety Disorder scale-7 (GAD-7) is a screening, diagnostic and monitoring tool that measures the severity of generalised anxiety in primary care settings (Spitzer, Kroenke, Williams, & Lowe, 2006). Each item is scored on a range from 0 (*not at all*) to 3 (*nearly every day*), with higher scores indicating higher levels of anxiety. The 7-item scale has high internal consistency (see Table 2) and good test-retest reliability ($r = 0.83$, Spitzer et al., 2006). The scale rates the frequency of symptoms, and a follow up, non-scored question assigns weight to the degree to which anxiety symptoms have affected a patient's level of functioning. This item was not completed at this study's participating sites.

The United States of America validation study demonstrated good sensitivity and specificity for GAD (Spitzer et al., 2006). Subsequent studies also reported generally good specificity and a range of sensitivities across samples (Table 2; García-Campayo et al., 2010; Kroenke, Spitzer, Williams, Monahan, & Löwe, 2007; Simpson, Glazer, Michalski, Steiner, & Frey, 2014; Zhong et al., 2015). Further reports indicated high sensitivity and good specificity for also detecting panic disorder, social anxiety disorder and PTSD (Kroenke et al., 2007). An optimal cut-point for any anxiety disorder was established as ≥ 9 (Kroenke et al., 2007) and ≥ 10 for GAD in Western samples (García-Campayo et al., 2010; Kroenke et al., 2007; Spitzer et al., 2006). Optimal cut-points for practical use in SA have not yet been established.

Studies across cultures and locations have substantiated the unidimensional structure of the GAD-7 (see Table 3; Bezuidenhout, 2018; García-Campayo et al., 2010; Henn & Morgan, 2019; Hinz et al., 2017; Jordan, Shedden-Mora, & Löwe, 2017; Löwe et al., 2008; Omani-Samani, Maroufizadeh, Ghaheri, & Navid, 2018; Zhong et el., 2015) and its factorial invariance for gender and age (Hinz et al., 2017; Löwe et al., 2008) with women reporting higher scores than men (Löwe et al., 2008). Evidence supporting construct and criterion validity (Barthel, Barkmann, Ehrhardt, Bindt, & International CDS Study Group, 2014; Bezuidenhout, 2018; García-Campayo et al., 2010; Omani-Samani et al., 2018) and cross-cultural validity (Zhong et al., 2015) has been reported. Initial SA validation on non-clinical samples found evidence of discriminant validity, acceptable Cronbach's alpha and concluded that the GAD-7 showed promise towards measurement fairness in SA (Bezuidenhout, 2018; Henn & Morgan, 2019).

Primary care post-traumatic stress disorder screen for DSM-5

Post-traumatic stress disorder is a syndrome that develops subsequent to exposure to traumatic events where an individual believed that there was a threat to life or physical integrity and safety and is characterised by a range of symptom clusters (APA, 2021). The primary care post-traumatic stress disorder screen for DSM-5 (PC-PTSD-5) was developed as a brief screen for PTSD in primary care settings using updated DSM-5 criteria. The 5-item screen enquires about the presence or absence of core PTSD symptoms, namely intrusive memory, avoidance, alterations in cognition

and mood and alternations in arousal and reactivity. The scale has high internal consistency (Table 2; Bovin et al., 2021; Jung et al., 2018), with good test-retest reliability (r = 0.89) and concurrent validity reported (Jung et al., 2018). The PC-PTSD-5 has demonstrated excellent diagnostic accuracy (see Table 2), with a cut-point of ≥ 3 offering optimal sensitivity and specificity (Bovin et al., 2021; Jung et al., 2018; Prins et al., 2016). The only report of previous use in SA that could be located (with emergency medical personnel; Van Wijk et al., 2020) did not report psychometric data.

Intense (panic-like) anxiety

Panic disorder is a syndrome characterised by repeated episodes of sudden onset intense apprehension and fearfulness in the absence of actual danger, accompanied by a range of discomforting physical symptoms (APA, 2021). The 2-item scale for panic-like anxiety came from the Guide for Aviation Medical Examiners (SACAA, 2017; the third CAA item – seeking urgent medical advice because of anxiety – was not included). The 2-item scale focuses on the sudden and intense experience of anxiety symptoms, as well as unexplained physical sensations associated with anxiety. A *YES* answer to either item would result in referral for further assessment. It was included based on its current use in industry, although no studies of its usefulness could be located.

CAGE scale

Alcohol use disorder is a catch-all diagnosis encompassing varying degrees of excessive use of alcohol (including abuse and dependence) (APA, 2021). Problematic alcohol use was determined using the 4-item CAGE (Ewing, 1984). The CAGE questionnaire has been extensively evaluated for use in identifying alcoholism and is considered a validated screening technique (cf. Dhalla & Kopec, 2007, for review). High sensitivity and specificity were reported for the identification of excessive, that is, problem, drinking, as well as for the identification of alcoholism (Table 2; Claassen, 1999; see also Williams, 2014, for review). High test-retest reliability has also been described (r > 0.80; Dhalla & Kopec, 2007). Historically, various cut-scores have been proposed based on different demographic factors (e.g. gender), and currently a score of ≥ 2 is generally considered indicative for concern (i.e. for alcohol dependence; Dhalla & Kopec, 2007; O'Brien, 2008; Vissoci et al., 2018; Williams, 2014). Studies from sub-Saharan Africa (Table 2 and Table 3) suggested good diagnostic utility (Claassen, 1999) and good internal reliability and unidimensional structure (Vissoci et al., 2018). In spite of its widespread use in SA (Labadarios, 2018), no reports on local validation of cut-points for the English version could be located.

Procedure

The measures were collated into booklet form, in the order presented here. Two measures, namely the BSI-18-S and the panic-like anxiety screen, were discontinued before the end of the study period and were removed from subsequent booklets. Each scale was presented in English, using the standard format and administration described in the respective source materials (e.g. manuals).

Each participant also partook in a clinical interview. This was conducted by clinical psychologists, who assessed – using DSM-5 criteria – the presence of disorders of mood (i.e. MDD), anxiety (i.e. GAD, panic disorder, PTSD), or substances (i.e. AUD). The assessment focussed on the specific syndromes listed here, and was therefore not inclusive of all presentations of poor mental health. Other conditions were identified and noted but not included in this study. Despite extensively reported criticisms (cf. Lynch, 2018, for overview), the DSM-5 (APA, 2013) remains the gold standard for clinical diagnostic purposes. In contrast to initial concerns (Chmielewski, Clark, Bagby, & Watson, 2015), excellent inter-rater interview-based diagnostic reliability (kappa > 0.70) has been reported for experienced psychiatrists and psychologists (Osório et al., 2019). The psychologists involved in the present study had at least 5 years experience in the occupational health surveillance context. The purpose of the interview assessment was to act as the reference standard (i.e. criterion measure) against which to evaluate the clinical utility of the brief mental health scales. Interviews took place within 24 hours of completing the screening booklet, and participants were allowed time off work to attend the interview. This study was incorporated into an ongoing occupational health screening programme, and the scales and psychological interview was administered as part of an annual occupational mental health review. Responses were entered into a spreadsheet, coded where appropriate and then irreversibly anonymised.

To ensure consistency of data gathering, three study review points were planned, the first at 300 cases, the second at 750 cases and the third at 1500 cases. Participating psychologists were further encouraged to share their clinical impressions and other concerns during monthly group supervision meetings. One purpose of the review was to consider the clinical usefulness of the scales, and if they were deemed to contribute little to the process or created an undue burden on the clinicians or participants, then to be discontinued. As a result of the health provision focus of the screening programme, interpretation of clinical utility was skewed to its practical impact more than on its psychometric characteristics. As mentioned earlier, two scales were discontinued early, and the available data for them will be reported under the Results section. All 1816 participants completed the remaining four scales, whilst only some completed the additional two scales prior to their discontinuation.

Statistical analysis

All statistical analyses, with the exclusion of the confirmatory factor analysis (CFA), were conducted with SPSS (version 27). Internal consistencies of the scales were examined with Cronbach's alpha, item-intercorrelations and corrected item-total correlations. Mplus 8.6 was used in both CFAs to assess unidimensionality and multigroup measurement invariance (Muthén & Muthén, 2017).

Dimensionality of the PHQ-9, GAD-7, PC-PTSD-5 and CAGE were examined with CFA. It was expected that the four scales will exhibit unidimensionality, that is, items or indicators loading highly on one latent factor each. All items were examined for distribution properties and deviation from normality. Skewness and non-normality influence the type of estimator used in the CFA. Usually, maximum likelihood (ML) is used, but for skew and non-normal data the estimators need to be robust and the choice depends amongst others on the nature of the indicators (Brown, 2015). Thus, for continuous variables (PHQ-9, GAD-7), the maximum likelihood - robust (MLR) estimator was used and for categorical responses (PC-PTSD-5, CAGE), weighted least squares - mean and variance-adjusted (WLSMV) (Muthén & Muthén, 2017). The global fit χ^2 would be preferred to be small and non-significant. Although this is rarely achieved, the following indices with cut points were taken into consideration. The standardised root mean square residual (SRMR) with good fit indicated by < 0.08 (Schreiber, Nora, Stage, Barlow, & King, 2006). The root mean square error of approximation (RMSEA) should be < 0.06 to < 0.08 for continuous data and < 0.06 for categorical data (Schreiber et al., 2006). Both the comparative fit index (CFI) and the Tucker–Lewis index (TLI) should be > 0.95 (Schreiber et al., 2006).

Local indications of misfit are the size of the standardised residuals and modification indices > 4 (Brown, 2015; Hair, Black, Babin, & Anderson, 2019). Usual indications of local problems on the models are standardised factor correlations out of range, negative error and factor variances, the significance of factor loadings, the size of parameter estimates and the reliability of indicators indicated by percentage of variance accounted for by the latent factors (indicated by the *R*-square of indicators). Modification indices should indicate no covariance between error variances, in this case referred to as within-construct error covariance (Hair et al., 2019).

Measurement invariance is a crucial aspect to assess for scales, especially if scores need to be compared across groups, whether they are language, gender or multicultural groups. Researchers often compare groups on test scores without considering measurement invariance (Brown, 2015). Scales need to be invariant with respect to the way the latent constructs are formed (configural invariance), the indicators or items should load similarly on latent factors across the groups (metric invariance) and lastly the origin of an indicator should be the same across groups, that is, they should have slopes (metric invariance) and similar origins on the y-axis (Wang & Wang, 2020). Testing for intercept invariance is called scalar equivalence. Thus, the process with testing for measurement invariance is to, firstly, look at the performance of a model in each subgroup sample (single group solutions) (see Table 4). Modifications to models may be made at this stage but if the groups' models differ in terms of specifications, one would be testing for partial measurement invariance (Byrne, 2012). Secondly, both groups are tested for factor structure (configural invariance), then for metric and scalar invariance. It is a hierarchical process thus one cannot proceed to nested models if model fit for the previous level fails (Kline, 2016). If modifications to the models can be substantiated, then the next level will be tested for partial measurement invariance given the restrictions placed on the model (Byrne, 2012).

The requirement for invariance is that the difference in global χ^2 between hierarchical models is not significant. In the case of the estimators used in this study, namely MLR for continuous indicators and WLSMV for binary indicators, the Satorra–Bentler correction for the difference between successive models were calculated because of differences not following a χ^2 distribution (Kline, 2016; Muthén & Muthén, 2017).

The measurement invariance for the PHQ-9, GAD-7, PC-PTSD-5 and the CAGE were evaluated first for gender (men and women; see Table 4) and then language (English first language speakers, and English second language speakers; Table 4). In each instance the group model results were provided as singular group solutions, and then in the order of configural, metric and scalar invariance. The measurement invariance of the PC-PTSD-5 and the CAG included only configural and scalar invariance because of the binary or categorical nature of their responses (Brown, 2015).

Criterion validity (and for the purpose of this study, also clinical utility) was explored through receiver operating/operator characteristics (ROC) curve analyses and positive and negative predictive values were calculated. Sensitivity and specificity data were calculated to address optimal cut-points for use in clinical practice. Receiver operating characteristics analysis is used to evaluate diagnostic tests and predictive models by plotting sensitivity versus specificity of a classification test, expressed as area under the curve (AUC). An AUC ≥ 0.70 is considered fair, ≥ 0.80 considered good, and ≥ 0.90 excellent (Safari, Baratloo, Elfil, & Negida, 2016). Sensitivity refers to the ability of a test to correctly identify persons with a condition, whilst specificity refers to the ability of a test to correctly identify people without the condition. Positive predictive value is the probability that persons with a positive screening test truly have the condition, whilst negative predictive value refers to the probability that persons with a negative screening test truly don't have the condition.

After measurement invariance was examined, socio-demographic effects were further explored using Pearson's correlation coefficients (for age effects) and *t*-tests for independent samples (for gender and language effects). Age and gender effects were previously reported (as discussed here), and this analysis served to explore whether different interpretative values (e.g. cut-points) might be required for different groups. Psychological scales often contain abstract concepts and in this sample

TABLE 4: Measurement invariance statistics for gender and language.

Gender measurement invariance	χ^2	*df*	$\Delta\chi^2$	Δdf	CFI	TLI	SRMR	RMSEA	90%CI
PHQ-9 single group solutions									
Men (*n* = 1318)	77.317***	27	-	-	0.949	0.932	0.039	0.038	0.028–0.048
Women (*n* = 498)	90.952***	27	-	-	0.907	0.876	0.046	0.069	0.054–0.085
PHQ-9 measurement invariance									
Configural	164.901***	54	-	-	0.936	0.914	0.041	0.048	0.039–0.056
Metric	165.440***	62	7.646	8	0.940	0.930	0.048	0.043	0.035–0.051
Scalar	188.583***	70	26.079***	8	0.931	0.929	0.050	0.043	0.036–0.051
Scalar (D4)	180.397***	69	8.344	7	0.935	0.933	0.049	0.042	0.035–0.050
GAD-7 single group solutions									
Men (*n* = 1048)	52.012***	14	-	-	0.960	0.940	0.034	0.051	0.037–0.066
Women (*n* = 416)	19.964	14	-	-	0.990	0.985	0.024	0.032	0.000–0.061
GAD-7 measurement invariance									
Configural	75.161***	28	-	-	0.970	0.956	0.032	0.048	0.035–0.061
Metric	79.963***	34	6.119	6	0.971	0.964	0.040	0.043	0.031–0.055
Scalar	95.478***	40	17.997**	6	0.965	0.963	0.042	0.044	0.032–0.055
Scalar (GA5)	90.084***	39	7.125	5	0.968	0.966	0.040	0.042	0.031–0.054
PC-PTSD-5 single group solutions									
Men (*n* = 1318)	19.891**	5	-	-	0.980	0.961	0.063	0.048	0.027–0.070
Women (*n* = 498)	9.903	5	-	-	0.991	0.982	0.058	0.044	0.000–0.085
PC-PTSD-5 measurement invariance									
Configural	28.873**	10	-	-	0.986	0.971	0.062	0.046	0.027–0.065
Scalar	30.957**	13	3.043	3	0.986	0.979	0.064	0.039	0.021–0.057
CAGE single group solution									
Men (*n* = 1318)	5.251	2	-	-	0.999	0.996	0.040	0.035	0.000–0.073
Women (*n* = 498)	Not determined	-	-	-	-	-	-	-	-
CAGE measurement invariance									
Configural	Not tested	-	-	-	-	-	-	-	-
Scalar	Not tested	-	-	-	-	-	-	-	-
Language measurement invariance									
PHQ-9 single group solutions									
English first (*n* = 344)	62.775***	27	-	-	0.884	0.845	0.061	0.062	0.042–0.082
English second (*n* = 1472)	94.826***	27	-	-	0.948	0.930	0.037	0.041	0.032–0.051
PHQ-9 measurement invariance									
Configural	155.305***	54	-	-	0.934	0.912	0.043	0.045	0.037–0.054
Metric	151.813***	62	5.035	8	0.941	0.932	0.048	0.040	0.032–0.048
Scalar	167.937***	70	8.001	8	0.936	0.934	0.048	0.039	0.032–0.047
GAD-7 single group solutions									
English first (*n* = 290)	24.621*	14	-	-	0.972	0.959	0.036	0.051	0.012–0.084
English second (*n* = 1174)	52.436***	14	-	-	0.965	0.948	0.030	0.048	0.035–0.063
GAD-7 measurement invariance									
Configural	75.526***	28	-	-	0.969	0.953	0.031	0.048	0.035–0.061
Metric	85.807***	34	10.238	6	0.966	0.958	0.041	0.046	0.034–0.058
Scalar	105.682***	40	30.461***	6	0.957	0.955	0.042	0.0477	0.036–0.058
PC-PTSD-5 single group solutions									
English first (*n* = 344)	Not determined	-	-	-	-	-	-	-	-
English second (*n* = 1472)	29.378***	5	-	-	0.977	0.953	0.065	0.058	0.039–0.078
PC-PTSD-5 measurement invariance									
Configural	Not tested	-	-	-	-	-	-	-	-
Scalar	Not tested	-	-	-	-	-	-	-	-
CAGE single group solutions									
English first (*n* = 344)	Not determined	-	-	-	-	-	-	-	-
English second (*n* = 1472)	5.834	2	-	-	0.999	0.996	0.037	0.036	0.000–0.072
CAGE measurement invariance									
Configural	Not tested	-	-	-	-	-	-	-	-
Scalar	Not tested	-	-	-	-	-	-	-	-

χ^2, Chi-square; *df*, degrees of freedom; $\Delta\chi^2$ based on Satorra–Bentler correction for MLR and WLSMV estimation; Δdf, degrees of freedom change; CFI, comparative fit index; TLI, Tucker–Lewis Index; SRMR, standardised root mean square residual; RMSEA, root mean square error of approximation; PHQ-9, Patient Health Questionnaire - 9; GAD-7, Generalised Anxiety Disorder scale - 7; PC-PTSD-5, Primary care Post traumatic stress disorder screen - 5; 90%CI, 90% confidence interval.

*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

were administered in English to a multi-language population. To explore the fairness of the scales – particularly for screening purposes – across different home languages (but with at least Grade 9 English literacy), the sample was divided into two groups, namely English first language (18.9% of the sample) and Non-English first

language (81.1% of the sample) to facilitate additional analysis.

Ethical considerations

The study has been approved by Stellenbosch University's Health Research Ethics Committee (#N20/07/078). All participants (n = 1816) gave written informed consent to the process and researchers only had access to de-identified data for analysis.

Results

Indicators of scale dimensionality are reported in Table 3, indicators of measurement invariance analysis are given in Table 4 and socio-demographic effects and criterion validity markers are presented in Table 5. Detailed sensitivity and specificity figures are presented in Table 6. Across all measures, age correlated significantly with scores. All the age correlations were negative, with very small effect sizes. Cronbach's alphas are reported in Table 5 and in no case did alpha improve through the deletion of items.

Although none of the four models tested for unidimensionality obtained a non-significant χ^2, the values were not excessively high. All other fit indices exceed the cut-points provided earlier. The TLI (0.93) for PHQ-9 was an exception, but the CFI was close enough to 0.95. The RMSEA was sufficiently small (0.04–0.05) for models, except for PC-PTSD-5, which still reached the criterion of < 0.06. The SRMR was smaller than 0.06 for all models. It can be accepted that all models exhibited sufficient fit to be evaluated as unidimensional scales (see Table 3). The details per scale are presented here.

In terms of measurement invariance, Table 4 shows that the single group solutions for men and women did not obtain a non-significant global χ^2 although SRMR were smaller than 0.08 and RMSEA was sufficiently low. Both CFI and TLI ranged in the region of 0.09 and it seems as if the smaller women sample fit the model less well than the model for men. The details per scale are presented here. Detail for the measurement invariance process for language for the four instruments are also presented in Table 4 and detailed here.

Primary Health Questionnaire-9

Acceptable Cronbach's alpha (Table 5) and corrected item-total correlations (Figure 1) were found, with inter-item correlations ranging from 0.22 to 0.60. During the CFA, the Primary Health Questionnaire-9 (PHQ-9) showed two modification indices higher than 20 for covariance between indicator error variances. Only substantial reasons would allow including these two within-construct error covariance to be freed for estimation (Hair et al., 2019). The content of the items, although somewhat related, would not warrant such a decision (Byrne, 2012). Standardised loadings were relatively uniform and high, ranging from 0.58 to 0.76 with Item 9 at 0.47. The scale demonstrated significant parameters (Table 3), low error, high communality

TABLE 5: Psychometric properties of study scales.

Measure	Scale construction			Socio-demographic effects						Clinical utility					
				Age		Gender		Language		ROC					
	M (s.d.)	Items	Alpha (95% CI)	*r*	*p*	Group difference	*p*	Group difference	*p*	AUC	95 % CI	PPV	NPV	SENS	SPEC
Patient Health Questionanaire-9	1.92 (3.35)	9	0.848							0.995	0.992–0.999	70.8	99.7	91.1	98.8
Total score			0.837–0.858	-0.189	< 0.001	t = -3.565	< 0.001	t = -1.787	0.074						
Cases (≥ 10)						χ^2 = 4.952	0.026	χ^2 = 0.253	0.615						
Brief Symptom Inventory 18 – somatisation scale		6	0.577												
Total score			0.522–0.627	-0.137	0.010	t = -5.432	< 0.001	t = -0.913	0.364						
Generalised Anxiety Disorder-7	1.56 (3.02	7	0.889							0.998	0.997–1.000	56.8	100	100	97.7
Total score			0.881–0.898	-0.165	< 0.001	t = -2.062	0.040	t = -0.734	0.464						
Cases (≥ 9)						χ^2 = 1.590	0.207	χ^2 = 0.055	0.815						
Primary Care Post-Traumatic Stress Disorder screen for DSM-5	0.17 (0.61)	5	0.696							0.989	0.982–0.996	46.1	99.9	86.7	98.7
Total score			0.674–0.718	-0.115	< 0.001	t = -1.808	0.071	t = -1.389	0.165						
Cases (≥ 3)						χ^2 = 3.744	0.053	χ^2 = 1.467	0.226						
Panic like anxiety		2	0.546												
Total score			0.476–0.607	-0.083	0.024	t = -2.189	0.029	t = 0.040	0.968						
Cases (≥ 1)						χ^2 = 3.364	0.067	χ^2 = 0.145	0.704						
CAGE scale	0.24 (0.66)	4	0.649							0.891	0.886–0.896	24.2	100	100	74.6
Total score			0.624–0.674	-0.138	< 0.001	t = 6.375	< 0.001	t = -4.619	< 0.001						
Cases (≥ 2)						χ^2 = 28.770	< 0.001	χ^2 = 16.282	< 0.001						

ROC, receiver operating/operator characteristic; AUC, area under the curve; 95% CI, 95% confidence interval; PPV, positive predictive value; NPV, negative predictive value; SENS, sensitivity; SPEC, specificity; s.d., standard deviation.

TABLE 6: Sensitivity and specificity indicators for PHQ-9, GAD-7, PC-PTSD-5, and CAGE.

Score	Sensitivity	Specificity
PHQ-9		
8	100	84.8
9	94.6	92.9
10	94.6	98.9
11	83.9	99.4
12	75.0	99.7
GAD-7		
7	100	96.8
8	100	97.7
9	96.6	98.8
10	96.6	99.4
11	86.2	99.6
PC-PTSD-5		
1	100	87.4
2	93.3	86.7
3	90.1	98.7
4	40.0	99.6
5	13.3	99.7
CAGE		
1	100	67.4
2	100	74.6
3	73.9	99.6
4	12.9	99.7

PHQ-9, Patient Health Questionnaire - 9; GAD-7, Generalised Anxiety Disorder scale - 7; PC-PTSD-5, Primary care Post traumatic stress disorder screen - 5.

as indicated by R-Square values for all indicators, and loaded high on each latent factor, providing sufficient evidence for unidimensionality.

The PHQ-9 for men and women were configural and metric invariant ($\Delta\chi^2 = 7.7$, $\Delta df = 8$) but did not reach scalar invariance ($\Delta\chi^2 = 26.1$, $\Delta df = 8$, $p < 0.001$). However, the examination of the modification indices showed that Item 4 influenced invariance and allowed its intercept to be freely estimated and permitted the remainder of intercepts to remain equivalent ($\Delta\chi^2 = 8.3$, $\Delta df = 7$). Note that the amended scalar model was compared with the metric model. Thus, PHQ-9 achieved partial measurement invariance on the scalar level, for gender.

The models for English first language speakers and for English second language speakers showed adequate fit for the RMSEA and SRMR indices. The smaller group of English first language speakers showed a CFI = 0.884 and TLI = 0.845 whilst the larger group of second language speakers were above 0.9 for the same indices. The global χ^2 for both groups were significant ($p < 0.001$). Full measurement invariance was demonstrated for configural, metric ($\Delta\chi^2 = 5.0$, $\Delta df = 8$) and scalar levels ($\Delta\chi^2 = 8.0$, $\Delta df = 8$), for language.

The PHQ-9 correlated significantly with the GAD-7, PC-PTSD-5 and CAGE and there were also significant comorbidities between MDD and GAD, PTSD and AUD (Table 7).

Excellent AUC was found (Table 5) and optimal sensitivity and specificity were obtained around a cut-point of ≥ 10 (Table 6). No significant language effects were found but there was a significant gender effect, where women reported more severe mood symptoms (Cohen's $d = 0.18$; mean difference = 0.6) and more proportional cases were reported. Given the partial scalar invariance, small effect size and small mean difference, it did not appear practically useful to develop separate cut-points for women and men.

TABLE 7: Inter-scale correlations for PHQ-9, GAD-7, PC-PTSD-5, and CAGE.

Variable	GAD-7	PC-PTSD-5	CAGE
PHQ-9	$r = 0.841$*	$r = 0.495$*	$r = 0.321$*
GAD-7	-	$r = 0.531$*	$r = 0.329$*
PC-PTSD-5	-	-	$r = 0.238$*

PHQ-9, Patient Health Questionnaire - 9; GAD-7, Generalised Anxiety Disorder scale - 7; PC-PTSD-5, Primary care Post traumatic stress disorder screen - 5.
*, $p < 0.001$.

Brief Symptom Inventory-18-S

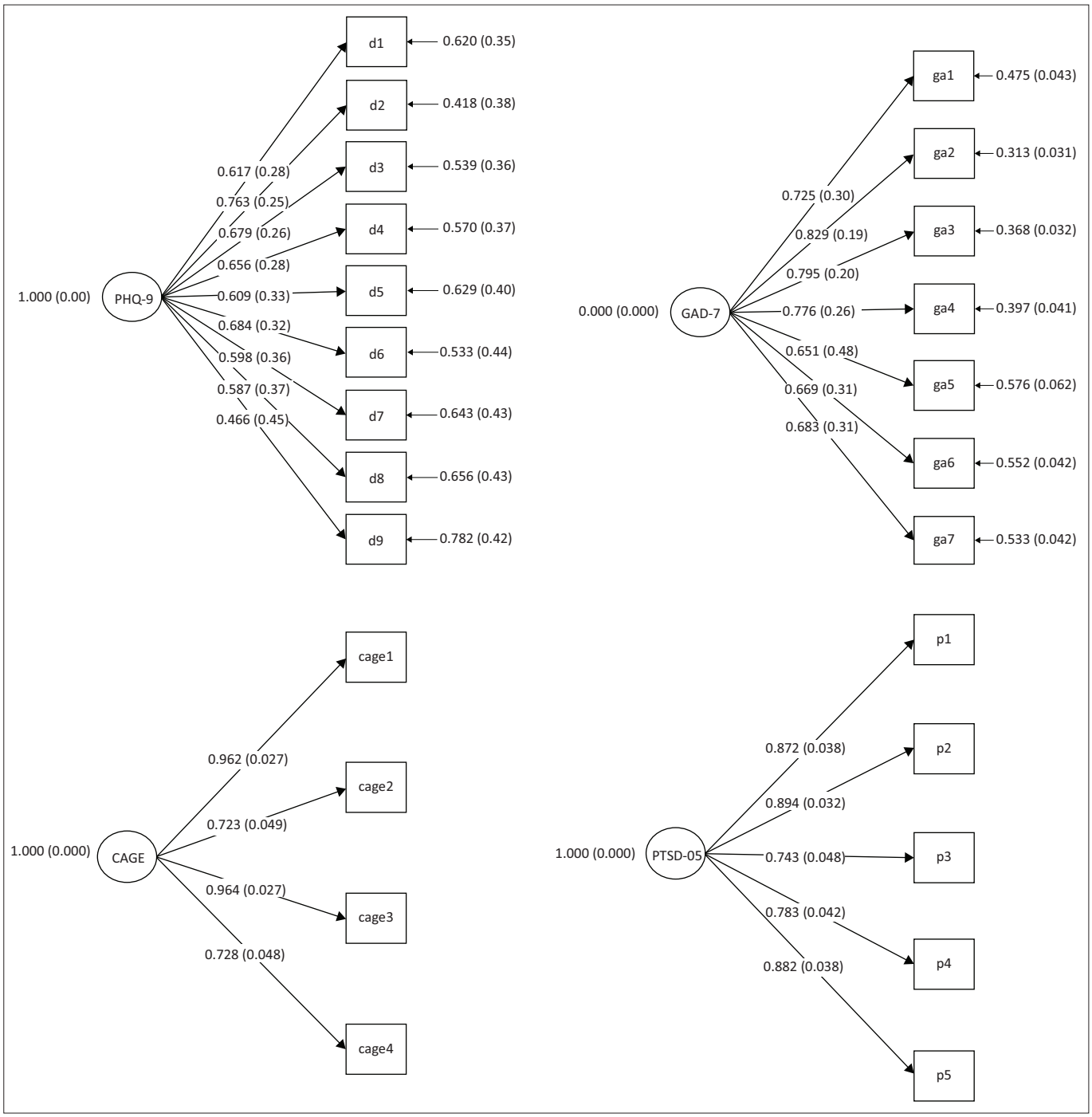
A progress review after 350 cases found little usefulness of this scale. There was a poor association with clinical outcomes, identifying only 50% of interview-determined cases of psychological distress (i.e. defined for this purpose as any DSM-5 disorder) and poor internal consistency. There were moderate correlations with other scales (PHQ-9: $r = 0.526$, $p < 0.001$; GAD-7: $r = 0.364$, $p < 0.001$), which all displayed better sensitivity and specificity. As a result of its poor clinical utility, its use was discontinued after 352 cases.

Generalised Anxiety Disorder scale-7

Acceptable Cronbach's alpha (Table 5) and corrected item-total correlations (Figure 1) were found, with inter-item correlations ranging from 0.46 to 0.72. During the CFA, the GAD-7 exhibited a similar situation as with the PHQ-9, with two high within-construct error covariances, but the same argument against freeing these parameters applied. The standardised loadings were consistently uniform and high, ranging from 0.65 to 0.83. The scale demonstrated significant parameters (Table 3), low error, high communality as indicated by R-Square values for all indicators and loaded high on each latent factor, providing sufficient evidence for unidimensionality.

The GAD-7 single group solutions showed that the model for women exhibited good fit with a non-significant χ^2 (19.964, $df = 14$) and all other fit indices well over the recommended limits for good fitting models. Except for the global χ^2 (52.0, $df = 14$, $p < 0.001$) the remainder of the fit indices also indicated a good fitting model for men. Similar results to the previous test were found with respect to measurement invariance: the instrument exhibited both configural and metric invariance but partial scalar invariance when the intercept for Item 5 was freely estimated ($\Delta\chi^2 = 7.1$, $\Delta df = 5$).

The GAD-7 showed adequate fit for both language groups for the CFI, TLI, RMSEA and SRMR indices. The global χ^2 for both groups were significant with the smaller English first language group significant at the 0.05 exceedance level and the larger group $p < 0.001$. The GAD-7 achieved configural



PHQ-9, Patient Health Questionnaire - 9; GAD-7, Generalised Anxiety Disorder scale - 7; PC-PTSD-5, Primary care Post traumatic stress disorder screen - 5;

FIGURE 1: Confirmatory factor analysis results for one factor models.

and metric invariance ($\Delta\chi^2$ = 10.238, Δdf = 6) but not scalar equivalence ($\Delta\chi^2$ = 30.4610, Δdf =6, $p < 0.001$). Modification indices showed no intercepts influencing the models thus partial invariance for intercepts was not possible.

The GAD-7 correlated significantly with the PHQ-9, PC-PTSD-5 and CAGE, and there were also significant comorbidities between GAD and MDD, PTSD and AUD disorder (Table 7).

Excellent AUC was found (Table 5) and good sensitivity and specificity were obtained around a cut-point of ≥ 9. In this sample, specificity was marginally improved (whilst maintaining sensitivity) when a score of ≥ 10 was used as cut-point (see Table 6). No significant language effects were found, but there was a significant gender effect, where women reported more severe anxiety symptoms (Cohen's *d* = 0.12; mean difference = 0.4). Given the absence of scalar invariance, small effect size and small mean difference, it did not appear practically useful to develop separate cut-points for women and men. The GAD-7 cases also included all cases of panic disorder and most cases of PTSD and were thus possibly more indicative of 'any' anxiety disorder than GAD only.

Primary care post-traumatic stress disorder screen for DSM-5

Cronbach's alpha was acceptable for research, but only borderline sufficient for clinical use (Table 5). Acceptable corrected item-total correlations (Figure 1) were found and inter-item correlations ranged from 0.22 to 0.50. During the CFA, the PC-PTSD-5 had no modification indices above 4 and standardised loadings ranged from 0.74 to 0.89. The scale demonstrated significant parameters (Table 3), low error, high communality as indicated by R-Square values for all indicators, and loaded high on each latent factor, providing sufficient evidence for unidimensionality.

The PC-PTSD-5 models for both men and women showed good fit indices with CFI, TLI, RMSEA and SRMR well within the limits for good fitting models. The model for women achieved a non-significant global χ^2 (9.903, df = 5). The instrument achieved both configural and scalar invariance ($\Delta\chi^2 = 3.0$, $\Delta df = 3$) for gender.

The model fit for PC-PTSD-5 first language speakers could not be determined because the residual covariance matrix (theta) is not positive definite and involved indicator Item 5. The global χ^2 for English second language speakers was significant ($\chi^2 = 29.378$, $df = 5$, $p < 0.001$). The second language single model CFI, TLI, RMSEA and SRMR indices were within acceptable limits. As a result of the undefined R-Square for indicator Item 5 in the English first language speaking group measurement invariance could not be evaluated for the PC-PTSD-5 for language.

The PC-PTSD-5 correlated significantly with the PHQ-9, GAD-7 and CAGE, and there were also significant comorbidities between PTSD and MDD and GAD (Table 6). Excellent AUC was reported (Table 5), and optimal sensitivity and specificity were obtained around a cut-point of ≥ 3 (see Table 6). There were no significant mean difference gender or language effects observed.

Panic-like anxiety

Early feedback from participating psychologists indicated scepticism regarding the usefulness of this scale, and after a progress review of the first 746 cases, it was discontinued because of poor specificity. Participants reported intense anxiety more often than what could be clinically diagnosed, with less than 40% of *YES* responses (to either item) associated with any actual diagnosis. All interview-confirmed panic cases were also identified through the GAD-7. Feedback from participating psychologists suggested that the high rate of false positives was more an indicator of non-pathological general psychological distress, rather than reflective of actual panic-like experiences.

CAGE

Cronbach's alpha was acceptable for research but not sufficient for clinical use (Table 5). Acceptable corrected item-total correlations (Figure 1) were found and inter-item correlations ranged from 0.35 to 0.63. During the CFA, the CAGE had no modification indices above 4. The standardised loadings were above 0.7 and topped out at above 0.9. The scale demonstrated significant parameters (Table 3), low error, high communality as indicated by R-Square values for all indicators and loaded high on each latent factor, providing sufficient evidence for unidimensionality.

The single group model for men fit extremely well with a non-significant χ^2 (5.251, df = 2) but the model for women could not be determined because the residual covariance matrix (theta) is not positive definite and involved indicator Item 3. Thus, measurement invariance for the CAGE could not be determined as a result of the undefined R-Square for indicator Item 3 in the women's group.

Again, the model fit for the English first language speakers could not be determined because the residual covariance matrix (theta) is not positive definite and involved indicator Item 3. The larger group, namely the English second language speakers single model yielded extremely good fit ($\chi^2 = 5.834$, df = 2, $p > 0.05$) and the CFI, TLI, RMSEA and SRMR indices very good fit. The measurement invariance for the CAGE could not be determined as a result of the undefined R-Square for indicator Item 3 in the English first language speaking group.

The CAGE correlated significantly with the PHQ-9, GAD-7 and PC-PTSD-5, and there were also significant comorbidities between AUD and MDD and GAD (Table 7).

Good AUC was reported and highest sensitivity and specificity were obtained around a cut-point of ≥ 2 (Table 5). There was a significant gender effect, where men reported more indicators of problematic alcohol use (Cohen's d = 0.39; mean difference = 0.2) and more proportional cases. There was also a significant language effect, where non-English first language speakers reported more indicators of problematic alcohol use (Cohen's d = 0.32; mean difference = 0.2). Given that measurement invariance for gender and language could not be determined, combined with the negligible effect sizes and small mean differences, it did not appear practically useful to develop separate cut-points based on gender or language.

Discussion

Outline of the results

Measures for clinical consideration: evidence of validity and practical implications for occupational health screening

The first objective was to provide evidence of structural validity. In this regard evidence of validity, based on internal structure, were found for all four scales. All four scales provided sufficient evidence for unidimensionality. Various degrees of measurement invariance were observed, with the two scales with binary responses not allowing

TABLE 8: Diagnosed comorbidities for MDD, GAD, PTSD, and AUD diagnosed comorbidities.

Variable	GAD	PTSD	AUD
MDD	$\chi^2 = 477.466$*	$\chi^2 = 204.600$*	$\chi^2 = 71.055$*
GAD	-	$\chi^2 = 260.739$*	$\chi^2 = 63.512$*
PTSD	-	-	$\chi^2 = 2.217$

MDD, major depressive disorder; GAD, Generalised Anxiety Disorder; PTSD, post-traumatic stress disorder; AUD, alcohol use disorders.
*, $p < 0.001$.

for full measurement invariance to be determined. Item-intercorrelations were generally acceptable and all inter-total correlations exceeded 0.3. The PHQ-9 and GAD-7 had good internal reliability, the PC-PTSD-5 was acceptable and the CAGE alpha coefficient, whilst acceptable for research, was questionable for clinical use. Furthermore, evidence of validity based on relationships to other variables were demonstrated in the expected significant correlations between the four scales (Table 7 and Table 8).

The second objective was to provide evidence of criterion validity and clinical utility. Evidence of validity based on test–criterion relationships were demonstrated through strong associations between scale scores and interview outcomes (as the references standard; Table 5). Furthermore, the PHQ-9, GAD-7 and PC-PTSD-5 displayed excellent screening accuracy, and the CAGE good accuracy, in this setting. The high screening accuracy may in part be a sampling artefact, as the participants were drawn from organisations with more ingrained systems and cultures that educate, promote and screen for mental illness. Such organisations tend to have workforce populations with higher rates of mental health literacy, who are more adept at recognising and reporting mental ill-health (Lieberman, 2019). The four scales further reported high negative predictive value, suggesting low rates of missed identification of risk – a desirable characteristic for screening tools. The variable and poorer positive predictive value was likely because of the relatively low prevalence of CMD in this generally healthy sample (Ranganathan & Aggarwal, 2018).

On a practical level, the PHQ-9 results supported previous recommendations that scores ≥ 10 be considered as a positive screen for depression in low- and middle-income contexts, as well as in occupational health settings (Akena et al., 2012; Volker et al., 2016). For the GAD-7, marginally better specificity was obtained when the cut-point for any anxiety disorder was raised to ≥ 10 (without sacrificing sensitivity). Furthermore, the GAD-7 appeared useful (in retrospect) to identify not only GAD but also panic and possibly even PTSD, which supported earlier international experience (Kroenke et al., 2007). Additional work will be necessary to determine whether different cut-points would be required for different presentations of disordered anxiety (e.g. GAD, panic disorder, etc.). The results of both scales supported previous reports on the higher mean scores of women compared with men (e.g. Löwe et al., 2008), although in neither case did the data require development of separate thresholds for women and men, which would simplify future interpretation during screening. For both the PHQ-9 and GAD-7 at least partial scalar invariance was observed for language, making these scales potentially useful for administration in multilingual workgroups.

The PC-PTSD-5 results supported previous recommendations that scores ≥ 3 be considered as a positive screen for further referral (Bovin et al., 2021; Jung et al., 2018; Prins et al., 2016). The slightly higher specificity than previous reports (Prins et al., 2016), together with the lack of significant gender and language effects (whilst acknowledging skewed subsample sizes) were encouraging for practical application. In a country with highly reported community level traumatic exposures and associated prevalence of PTSD (Edwards, 2005; Kaminer & Eagle, 2010; Peltzer & Pengpid, 2019), a screener as brief is this one may be particularly valuable. However, given that the interview-determined PTSD prevalence was only 0.8%, follow-up studies will be required to confirm the diagnostic accuracy of the PC-PTSD-5 in samples with higher prevalence of traumatic exposure or diagnosed PTSD. Follow-up studies would also be required to further explore measurement invariance across language groups.

The CAGE cut-point of ≥ 2 appear consistent with findings in sub-Saharan Africa studies (Claassen, 1999; Vissoci et al., 2018). However, in spite of the good screening accuracy reported, the CAGE might be somewhat less useful in the current context. The poor Cronbach's alpha may be questionable for clinical use, and the poorer specificity and PPV may lower screening efficiency by identifying too many false positive cases for referral. The gender effect found in this sample (i.e. higher mean scores of men compared with women), although small, is consistent with general reports, where the need for different cut-points for women and men has been exhaustively debated (cf. Dhalla & Kopec, 2007, for review).

The significant language effect observed in this sample poses a substantial challenge to the practical use of the CAGE, particularly as measurement invariance could not be determined. Non-English first language speakers reported more indicators of problematic alcohol use, although the mean score difference and effect size were small. It could be hypothesised that this sample of educated employees may have had better English proficiency than the general population (hence the small mean score difference), and that another sample may have greater difficulty with the language employed in the four items. All four items require some semantic interpretation and language background could therefore influence the reporting of the CAGE indicators of problematic alcohol use. Possible language effects may make the CAGE less suitable for use in the current SA multilingual context.

High levels of AUD have consistently been reported in SA society (Herman et al., 2009) and there remains a need for a

scale with acceptable diagnostic accuracy. In this regard, a scale like the AUDIT (Barbor, Higgens-Biddle, Saunders, & Monteiro, 2001; Saunders, Aasland, Babor, De la Fuente, & Grant, 1993), recommended by the World Health Organization, may be worth considering for local use.

Discontinued measures

The BSI-18-S was discontinued early because of questionable psychometric properties and poor clinical utility. It differentiated poorly between diagnostic cases, supporting the finding of Recklitis et al. (2017, p. 1197) who concluded that the BSI-18 should not be used as a stand-alone screening measure for making clinical decisions (i.e. referral for mental health follow-up). It could be hypothesised that the poor clinical utility was because of this educated sample exhibiting the necessary vocabulary to be more specific in reporting their distress (e.g. as mood or anxiety). Furthermore, a 6-item scale may not be sufficient to tap into a construct as complex as somatisation, especially as the BSI was initially designed for use with medical patients (Derogatis, 2001) and possibly was not a good fit in a population of generally healthy adults.

The 2-item panic-like anxiety scale was discontinued early because of poor specificity. This scale was adapted from civil aviation guidelines (SACAA, 2017), which were originally intended for a very specific group (e.g. aircrew), and may not be equally useful in a general industry population. This leaves two possible avenues for further consideration: Firstly, two items make a very brief scale to screen for a common and complex condition such as panic disorder. Future research may be usefully directed towards identifying additional items to improve its utility to screen for panic. Secondly, there may be no need for an additional intense panic-like measure in this context, as the GAD-7 in this sample also identified all panic cases. Thus, simply using the GAD-7 as general screening for multiple anxiety disorders might be sufficient, especially as all positive screening outcomes are automatically referred for further assessment.

Practical implications

The findings of this study point to an opportunity to more fully realise the strategic role that brief, locally validated and clinically efficacious screening measures can play in facilitating more efficient access points and referral pathways for mental health support in South African occupational healthcare. This is especially pertinent given the role that imprecise and non-normed screening measures play not only in the over or under-diagnosis of CMD but also the inappropriate allocation of, and expenditure on, intervention opportunities in what are often resource-limited occupational healthcare support systems. Alone, however, validated mental health screening measures do not adequately solve the burden of CMD in workplace settings and can in fact prove to be counterproductive should the 'point of screening' not be coupled to appropriate post-screening referral and treatment pathways (Joyce et al., 2016). Furthermore, whilst routine occupational mental health screening for CMD is evermore *en vogue* for its cost effectiveness (Dobson et al., 2018) and its role in establishing workplace cultures of 'continuous health promotion' (Magnavita, 2018), this does not eliminate longer standing critiques that such screenings heighten experiences of workplace stigma (Solomon, Mikulincer, & Flum, 1989). For this reason, the screening measures recommended here need to be embedded within workplace programmes of mental health education.

Limitations

The results from this convenience sample cannot be generalised to populations with lower formative education levels without further evidence of validity. Furthermore, the lack of comorbidity data – likely a significant proportion (Nel, Augustyn, & Bartman, 2018) – was not known for this data set. It is possible, and would require further investigation, that against the background of multiple comorbidities, scale scores could have reflected general mental distress, rather than specific diagnoses. The scales reported here measured the severity, or presence, of selected conditions, but not the extent of impact on daily life (PHQ-9 and GAD-7 have additional items to measure the degree to which symptoms have affected patients' level of functioning, but they were not available for this study). Future studies may be valuable in clarifying associations between severity scores and impact on level of functioning in local SA samples. Inter-rater reliability – when using multiple psychologists for DMS-5 based assessments – were not available for this study and may need to be accommodated in future protocols. Lastly, further research may require larger samples to investigate the measurement invariance of these scales across different socio-demographic variables.

Conclusion

This study reported evidence of structural and criterion validity for the four scales when administered in local occupational health surveillance settings. A particular benefit of the PHQ-9, GAD-7 and PC-PTSD-5 is that the same reference norms appear useful – for now – across gender and language backgrounds, at least in workplace populations with a minimum of 9 years of formal schooling. However, there remains a need for larger scale 'general population' studies to establish their utility in a more diverse range of occupational environments and workplaces, where systems and cultures of mental health promotion and intervention are less ingrained and practised.

For practical application, the PHQ-9, GAD-7 and PC-PTSD-5 demonstrated good diagnostic accuracy and – where there is a relatively highly educated and psychologically literate occupational sample – confirmed that targeted mental health screening presents potential clinical utility for identification, referral and intervention within occupational health surveillance infrastructure.

Acknowledgements

Competing interests

The authors declared that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Authors' contributions

C.H.v.W. and J.H.M. conceptualised the study. C.H.v.W., J.H.M. and D.J.F.M. contributed to the analysis of data and were involved in the final review and editing of the article.

Funding information

The authors received no financial support for the research, authorship, and/or publication of this article.

Data availability

The data that support the findings of this study are available from the corresponding author, C.v.W., upon reasonable request. The data are not publicly available due to privacy and ethical considerations.

Disclaimer

The views and opinions expressed in the article are those of the authors and do not necessarily reflect an official policy or position of any affiliated agency of the authors.

References

Adams, R.E., Boscarino, J.A., & Galea, S. (2006). Alcohol use, mental health status and psychological well-being 2 years after the World Trade Center attacks in New York City. *American Journal of Drug and Alcohol Abuse, 32*(2), 203–224. https://doi.org/10.1080%2F00952990500479522

Adewuya, A.O., Ola, B.A., & Afolabi, O.O. (2006). Validity of the patient health questionnaire (PHQ-9) as a screening tool for depression amongst Nigerian university students. *Journal of Affective Disorders, 96*(1–2), 89–93. https://doi.org/10.1016/j.jad.2006.05.021

Akena, D., Joska, J., Obuku, E.A., Amos, T., Musisi, S., & Stein, D.J. (2012). Comparing the accuracy of brief versus long depression screening instruments which have been validated in low- and middle-income countries: A systematic review. *BMC Psychiatry, 12*, Article 187. https://doi.org/10.1186/1471-244X-12-187

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Association.

American Psychological Association. (2021). *APA dictionary*. Retrieved from https://dictionary.apa.org/alcohol-use-disorder

Asami, Y., Goren, A., & Okumura, Y. (2015). Work productivity loss with depression, diagnosed and undiagnosed, among workers in an internet-based survey conducted in Japan. *Journal of Occupational and Environmental Medicine, 57*(1), 105–110. https://doi.org/10.1097/jom.0000000000000310

Asner-Self, K.K., Schreiber, J.B., & Marotta, S.A. (2006). A cross-cultural analysis of the Brief Symptom Inventory-18. *Cultural Diversity and Ethnic Minority Psychology, 12*(2), 367–375. https://doi.org/10.1037/1099-9809.12.2.367

Barbor, T.F., Higgens-Biddle, J.C., Saunders, J.B., & Monteiro, M.G. (2001). AUDIT. In *The alcohol use disorders identification test. Guidelines for use in primary health care* (2nd ed.). World Health Organization. Retrieved from https://www.who.int/publications/i/item/audit-the-alcohol-use-disorders-identification-test-guidelines-for-use-in-primary-health-care

Barthel, D., Barkmann, C., Ehrhardt, S., Bindt, C., & International CDS Study Group. (2014). Psychometric properties of the 7-item Generalized Anxiety Disorder scale in antepartum women from Ghana and Côte d'Ivoire. *Journal of Affective Disorders, 169*, 203–211. https://doi.org/10.1016/j.jad.2014.08.004

Bezuidenhout, D. (2018). *Validation of the general anxiety disorder – 7 in a non-clinical sample of South African employees* (Unpublished master's thesis). University of Johannesburg. Retrieved from http://hdl.handle.net/10210/402688

Bhana, A., Rathod, S.D., Selohilwe, O., Kathree, T., & Petersen, I. (2015). The validity of the Patient Health Questionnaire for screening depression in chronic care patients in primary health care in South Africa. *BMC Psychiatry, 15*, Article 118. https://doi.org/10.1186/s12888-015-0503-0

Botha, M.N. (2011). *Validation of the Patient Health Questionnaire (PHQ-9) in an African context*. Unpublished master's thesis. North-West University. Retrieved from http://hdl.handle.net/10394/4647

Bovin, M.J., Kimerling, R., Weathers, F.W., Prins, A., Marx, B.P., Post, E.P., & Schnurr, P.P. (2021). Diagnostic accuracy and acceptability of the primary care posttraumatic stress disorder screen for the diagnostic and statistical manual of mental disorders (5th ed.) among US veterans. *JAMA Network Open, 4*(2), e2036733. https://doi.org/10.1001/jamanetworkopen.2020.36733

Brown, T.A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: The Guilford Press.

Byrne, B.M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York: Routledge.

Carlson, L.E., Angen, M., Cullum, J., Goodey, E., Koopmans, J., Lamont, L., …Bultz, B.D. (2004). High levels of untreated distress and fatigue in cancer patients. *British Journal of Cancer, 90*, 2297–2304. https://doi.org/10.1038/sj.bjc.6601887

Chmielewski, M., Clark, L.A., Bagby, R.M., & Watson, D. (2015). Method matters: Understanding diagnostic reliability in DSM-IV and DSM-5. *Journal of Abnormal Psychology, 124*(3), 764–769. https://doi.org/10.1037/abn0000069

Cholera, R., Gaynes, B.N., Pence, B.W., Bassett, J., Qangule, N., Macphail, C., …Miller, W.C. (2014). Validity of the Patient Health Questionnaire-9 to screen for depression in a high-HIV burden primary healthcare clinic in Johannesburg, South Africa. *Journal of Affective Disorders, 167*, 160–166. https://doi.org/10.1016/j.jad.2014.06.003

Claassen, J.N. (1999). The benefits of the CAGE as a screening tool for alcoholism in a closed rural South African community. *South African Medical Journal, 89*(9), 976–979.

Derogatis, L.R. (2001). *BSI 18, Brief Symptom Inventory 18: Administration, scoring and procedure manual*. Minneapolis: NCS Pearson.

Dhalla, S., & Kopec, J.A. (2007). The CAGE questionnaire for alcohol misuse: A review of reliability and validity studies. *Clinical & Investigative Medicine, 30*(1), 33–41. https://doi.org/10.25011/cim.v30i1.447

Dobson, K.S., Szeto, A., Knaak, S., Krupa, T., Kirsh, B., Luong, D., …Pietrus, M. (2018). Mental health initiatives in the workplace: Models, methods and results from the Mental Health Commission of Canada. *World Psychiatry, 17*(3), 370–371. https://doi.org/10.1002%2Fwps.20574

Docrat, S., Besada, D., Cleary, S., Daviaud E., & Lund, C. (2019). Mental health system costs, resources and constraints in South Africa: A national survey. *Health Policy and Planning, 34*(9), 706–719. https://doi.org/10.1093/heapol/czz085

Draguns, J.G., & Tanaka-Matsumi, J. (2003). Assessment of psychopathology across and within cultures: Issues and findings. *Behaviour Research and Therapy, 41*(7), 755–776. https://doi.org/10.1016/s0005-7967(02)00190-0

Edwards, D. (2005). Post-traumatic stress disorder as a public health concern in South Africa. *Journal of Psychology in Africa, 15*(2), 125–134. https://doi.org/10.4314/jpa.v15i2.30650

Employment Equity Act. (1998). *Employment Equity Act 55 of 1998*. Retrieved from https://www.gov.za/documents/employment-equity-act

European Federation of Psychologists' Associations (2013). *EFPA review model for the description and evaluation of psychological and educational tests, version 4.2.6*. Retrieved from http://assessment.efpa.eu/documents-/

Ewing, J.A. (1984). Detecting Alcoholism: The CAGE Questionnaire. *JAMA, 252*, 1905–1907. https://doi.org/10.1001/jama.1984.03350140051025

Franke, G.H., Jaeger, S., Glaesmer, H., Barkmann, C., Petrowski, K., & Braehler, E. (2017). Psychometric analysis of the Brief Symptom Inventory 18 (BSI-18) in a representative German sample. *BMC Medical Research Methodology, 17*(1), 14. https://doi.org/10.1186/s12874-016-0283-3

García-Campayo, J., Zamorano, E., Ruiz, M.A., Pardo, A., Pérez-Páramo, M., López-Gómez, V., …Rejas, J. (2010). Cultural adaptation into Spanish of the generalized anxiety disorder-7 (GAD-7) scale as a screening tool. *Health and Quality of Life Outcomes, 8*, 1–11. https://doi.org/10.1186/1477-7525-8-8

Gilbody, S., Richards, D., & Barkham, M. (2007). Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ–9 and CORE–OM. *British Journal of General Practice, 57*, 650–652.

Hair, Jr., J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2019). *Multivariate data analysis* (8th ed.). Andover: Cengage.

Henn, C., & Morgan, B. (2019). Differential item functioning of the CESDR-R and GAD-7 in African and white working adults. *South Africa Journal of Industrial Psychology, 45*, Article a1663. https://doi.org/10.4102/sajip.v45i0.1663

Herman, A.A., Stein, D.J., Seedat, S., Heeringa, S.G., Moomal, H., & Williams, D.R. (2009). The South African Stress and Health (SASH) study: 12-month and lifetime prevalence of common mental disorders. *South African Medical Journal, 99*(5 Pt 2), 339–344.

Hilton, M.F., & Whiteford, H.A. (2010). Associations between psychological distress, workplace accidents, workplace failures and workplace successes. *International Archives of Occupational and Environmental Health, 83*(8), 923–933. https://doi.org/10.1007/s00420-010-0555-x

Hinz, A., Klein, A.M., Brähler, E., Glaesmer, H., Luck, T., Riedel-Heller, S.G., …Hilbert, A. (2017). Psychometric evaluation of the Generalized Anxiety Disorder Screener GAD-7, based on a large German general population sample. *Journal of Affective Disorders, 210*, 338–344. https://doi.org/10.1016/j.jad.2016.12.012

Jacob, N., & Coetzee, D. (2018). Mental illness in the Western Cape Province, South Africa: A review of the burden of disease and healthcare interventions. *South African Medical Journal, 108*(3), 176–180. https://doi.org/10.7196/samj.2018.v108i3.12904

Jain, G., Roy, A., Harikrishnan, V., Yu, S., Dabbous, O., & Lawrence, C. (2013). Patient-reported depression severity measured by the PHQ-9 and impact on work productivity: Results from a survey of full-time employees in the United States. *Journal of Occupational and Environmental Medicine, 55*(3), 252–258. https://doi.org/10.1097/jom.0b013e31828349c9

Jordan, P., Shedden-Mora, M.C., & Löwe, B. (2017). Psychometric analysis of the Generalized Anxiety Disorder scale (GAD-7) in primary care using modern item response theory. *PLoS One, 12*(8), e0182162. https://doi.org/10.1371/journal.pone.0182162

Joyce, S., Modini, M., Christensen, H., Mykletun, A., Bryant, R., Mitchell, P.B., & Harvey, S.B. (2016). Workplace interventions for common mental disorders: A systematic meta-review. *Psychological Medicine, 46*(4), 683–697. https://doi.org/10.1017/s0033291715002408

Jung, Y., Kim, D., Kim, W., Roh, D., Chae, J., & Park, J. (2018). A brief screening tool for PTSD: Validation of the Korean Version of the Primary Care PTSD Screen for DSM-5 (K-PC-PTSD-5). *Journal of Korean Medical Science, 33*(52), e338. https://doi.org/10.3346/jkms.2018.33.e338

Kaminer, D., & Eagle, G. (2010). *Traumatic stress in South Africa*. Johannesburg: Wits University Press.

Kessler, R.C., Lane, M., Stang, P.E., & Van Brunt, D.L. (2009). The prevalence and workplace costs of adult attention deficit hyperactivity disorder in a large manufacturing firm. *Psychological Medicine, 39*(1), 137–147. https://doi.org/10.1017/S0033291708003309

Kigozi, G. (2020). Confirmatory factor analysis of the Patient Health Questionnaire-9: A study amongst tuberculosis patients in the Free State province. *Southern African Journal of Infectious Diseases, 35*(1), 242. https://doi.org/10.4102/sajid.v35i1.242

Kline, R.B. (2016). Principles and practice of structural equation modelling (4th ed.). New York, NY: Guilford Publications.

Kroenke, K., Spitzer, R.L., & Williams, J.B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16(9)*, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Spitzer, R.L., Williams, J.B.W., Monahan, P.O., & Löwe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine, 146*, 317–325. https://doi.org/10.7326/0003-4819-146-5-200703060-00004

Labadarios, G. (2018). *Determination of a brief AUDIT screening questionnaire to identify women at risk of harmful and hazardous alcohol consumption in primary care settings*. Unpublished master's thesis. University of Cape Town. Retrieved from http://hdl.handle.net/11427/29356

Lamela, D., Soreira, C., Matos, P., & Morais, A. (2020). Systematic review of the factor structure and measurement invariance of the patient health questionnaire-9 (PHQ-9) and validation of the Portuguese version in community settings. *Journal of Affective Disorders, 276*, 220–233. https://doi.org/10.1016/j.jad.2020.06.066

Lieberman, C. (2019, August 14). What wellness programs don't do for workers. *Harvard Business Review*. Retrieved from https://hbr.org/2019/08/what-wellness-programs-dont-do-for-workers

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Yorck Herzberg, P. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care, 46*(3), 266–274. https://doi.org/10.1097/mlr.0b013e318160d093

Lynch, T. (2018). The validity of the DSM: An overview. *Irish Journal for Counselling and Psychology, 18*(2), 5–10.

Magnavita, N. (2018). Medical surveillance, continuous health promotion and a participatory intervention in a small company. *International Journal of Environmental Research and Public Health, 15*(4), 662. https://doi.org/10.3390/ijerph15040662

Mall, S., Lund, C., Vilagut, G., Alonso, J., Williams, D.R., & Stein, D.J. (2015). Days out of role due to mental and physical illness in the South African stress and health study. *Social Psychiatry and Psychiatric Epidemiology, 50*(3), 461–468. https://doi.org/10.1007/s00127-014-0941-x

Manea, L., Gilbody, S., & McMillan, D. (2012). Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): A meta-analysis. *Canadian Medical Association Journal, 184*(3), e191–e196. https://doi.org/10.1503/cmaj.110829

Maroufizadeh, S., Omani-Samani, R., Almasi-Hashiani, A., Amini, P., & Sepidarkish, M. (2019). The reliability and validity of the Patient Health Questionnaire-9 (PHQ-9) and PHQ-2 in patients with infertility. *Reproductive Health, 16*(1), 137. https://doi.org/10.1186/s12978-019-0802-x

Marx, B., Prins, A., Bovin, M.J., Weathers, F., Schnurr, P., Kaloupek, D., & Keane, T. (2014). *Performance of the PCL-5 and the PC-PTSD-5 relative to the CAPS-5 in diagnosing PTSD among veterans [Abstract]*. Paper presented at the 30th Annual Meeting of the International Society for Traumatic Stress Studies, Miami, FL, 6–8 November 2014. Retrieved from https://istss.org/ISTSS_Main/media/Documents/2014SessionAbstracts.pdf

McCann, M., Harker-Burnhams, N.H., Albertyn, C., & Bhoola, U. (2011). *Alcohol, drugs and employment* (2nd ed.). Claremont: Juta & Co Ltd.

Meachen, S.J., Hanks, R.A., Millis, S.R., & Rapport, L.J. (2008). The reliability and validity of the Brief Symptom Inventory-18 in persons with traumatic brain injury. *Archives of Physical Medical Rehabilitation, 89*(5), 958–965. https://doi.org/10.1016/j.apmr.2007.12.028

Mulvaney-Day, N., Marshall, T., Downey-Piscopo, K., Korsen, N., Lynch, S., Karnell, L.H. …Ghose, S.S. (2017). Screening for behavioral health conditions in primary care settings: A systematic review of the literature. *Journal of General Internal Medicine, 33*(3), 335–346. https://doi.org/10.1007/s11606-017-4181-0

Muthén, L.K., & Muthén, B.O. (2017). *Mplus user's guide* (8th ed.). Los Angeles: Muthén & Muthén.

Nel, C., Augustyn, L., & Bartman, N. (2018). Anxiety disorders: Psychiatric comorbidities and psychosocial stressors among adult outpatients. *South African Journal of Psychiatry, 24*, a1138. https://doi.org/10.4102/sajpsychiatry.v24i0.1138

Newcomb, R.D., Steffen, M.W., Breeher, L.E., Sturchio, G.M., Murad, M.H., Wang, Z., & Molella, R.G. (2016). Screening for depression in the occupational health setting. *Occupational Medicine, 66*(5), 390–393. https://doi.org/10.1093/occmed/kqw043

O'Brien, C.P. (2008). The CAGE questionnaire for detection of alcoholism: A remarkably useful but simple tool. *JAMA, 300*(17), 2054–2056. https://doi.org/10.1001/jama.2008.570

Omani-Samani, R., Maroufizadeh, S., Ghaheri, A., & Navid, B. (2018). Generalized anxiety Disorder-7 (GAD-7) in people with infertility: A reliability and validity study. *Middle East Fertility Society Journal, 23*(4), 446–449. https://doi.org/10.1016/j.mefs.2018.01.013

Osório, F.L., Loureiro, S.R., Hallak, J., Machado-de-Sousa, J.P., Ushirohira, J.M., Baes, C., …Crippa, J. (2019). Clinical validity and intrarater and test-retest reliability of the Structured Clinical Interview for DSM-5 – Clinician Version (SCID-5-CV). *Psychiatry and Clinical Neurosciences, 73*(12), 754–760. https://doi.org/10.1111/pcn.12931

Palmer, K.T., D'Angelo, S., Harris, E.C., Linaker, C., & Coggon, D. (2014). The role of mental health problems and common psychotropic drug treatments in accidental injury at work: A case–control study. *Occupational and Environmental Medicine, 71*(5), 308–312. https://doi.org/10.1136/oemed-2013-101948

Peltzer, K., & Pengpid, S. (2019). High physical activity is associated with post-traumatic stress disorder among individuals aged 15 years and older in South Africa. *South African Journal of Psychiatry, 25*, Article a1329. https://doi.org/10.4102/sajpsychiatry.v25i0.1329

Pence, B.W., Gaynes, B.N., Atashili, J., O'Donnell, J.K., Tayong, G., Kats, D., …Ndumbe, P.M. (2012). Validity of an interviewer-administered patient health questionnaire-9 to screen for depression in HIV-infected patients in Cameroon. *Journal of Affective Disorders, 143*(1–3), 208–213. https://doi.org/10.1016/j.jad.2012.05.056

Petkus, A.J., Gum, A.M., Small, B., Malcarne, V.L., Stein, M.B., & Wetherell, J.L. (2010). Evaluation of the factor structure and psychometric properties of the Brief Symptom Inventory-18 with homebound older adults. *International Journal of Geriatric Psychiatry, 25*(6), 578–587. https://doi.org/10.1002/gps.2377

Petrowski, K., Schmalbach, B., Jagla, M., Franke, G.H., & Brähler, E. (2018). Norm values and psychometric properties of the Brief Symptom Inventory-18 regarding individuals between the ages of 60 and 95. *BMC Medical Research Methodology, 18*(1), 164. https://doi.org/10.1186/s12874-018-0631-6

Prins, A., Bovin, M.J., Smolenski, D.J., Mark, B.P., Kimerling, R., Jenkins-Guarnieri, M.A. … Tiet, Q.Q. (2016). The primary care PTSD screen for DSM-5 (PC-PTSD-5): Development and evaluation within a Veteran primary care sample. *Journal of General Internal Medicine, 31*, 1206–1211. https://doi.org/10.1007/s11606-016-3703-5

Ranganathan, P., & Aggarwal, R. (2018). Common pitfalls in statistical analysis: Understanding the properties of diagnostic tests – Part 1. *Perspectives in Clinical Research, 9*(1), 40–43. https://doi.org/10.4103/picr.PICR_170_17

Recklitis, C.J., Blackmon, J.E., & Chang, G. (2017). Validity of the Brief Symptom Inventory-18 (BSI-18) for identifying depression and anxiety in young adult cancer survivors: Comparison with a structured clinical diagnostic interview. *Psychological Assessment, 29*(10), 1189–1200. https://doi.org/10.1037/pas0000427

Rossouw, L., Seedat, S., Emsley, R.A., Suliman, S., & Hagemeister, D. (2013). The prevalence of burnout and depression in medical doctors working in the Cape Town Metropolitan Municipality community healthcare clinics and district hospitals of the Provincial Government of the Western Cape: A cross-sectional study. *South African Family Practice, 55*(6), 567–573. https://doi.org/10.4102/phcfm.v10i1.1568

Saunders, J.B., Aasland, O.G., Babor, T.F., De la Fuente, J.R., & Grant, M. (1993). Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption – II. *Addiction, 88*(6), 791–803. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Safari, S., Baratloo, A., Elfil, M., & Negida, A. (2016). Evidence based emergency medicine; Part 5 receiver operating curve and area under the curve. *Archives of Academic Emergency Medicine, 4*(2), 111–113.

Schaap, P., & Kekana, E. (2016). The structural validity of the experience of work and life circumstances questionnaire (WLQ). *South African Journal of Industrial Psychology, 42*(1), a1349. https://doi.org/10.4102/sajip.v42i1.1349

Schoeman, R. (2017, August 31). Mental health problems cost SA's economy billions per year. *Financial Mail*. Retrieved from https://www.businesslive.co.za/fm/features/2017-08-31-mental-health-problems-cost-sas-economy-billions-per-year

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

Simpson, W., Glazer, M., Michalski, N., Steiner, M., & Frey, B.N. (2014). Comparative efficacy of the generalized anxiety disorder 7-item scale and the Edinburgh Postnatal Depression Scale as screening tools for generalized anxiety disorder in pregnancy and the postpartum period. *Canadian Journal of Psychiatry, 59*(8), 434–440. https://doi.org/10.1177/070674371405900806

Soares, S.M., Gelmini, S., Brandão, S.S.S., & Silva, J.M.C. (2018). Workplace accidents in Brazil: Analysis of physical and psychosocial stress and health-related factors. *Revista de Administração Mackenzie, 19*(3), Article eRAMG170131. https://doi.org/10.1590/1678-6971/eramg170131

Solomon, Z., Mikulincer, M., & Flum, H. (1989). The implications of life events and social integration in the course of combat-related post-traumatic stress disorder. *Social Psychiatry and Psychiatric Epidemiology, 24*, 41–48. https://doi.org/10.1007/BF01788199

South African Civil Aviation Authority. (2017). *Guide for aviation medical examiners* (revision 21 July 2017). Retrieved from http://www.caa.co.za/Documents/Aviation%20Medicine/Dames%20Guide.pdf

South African Police Service. (2016). SAPS employee health and wellness, presentation to the Parliamentary Portfolio Committee on Police, 2016 February 17. Retrieved from https://africacheck.org/wp-content/uploads/2016/02/EHW-presentation-PCOP-17-Feb-2016-1.pdf

Spitzer, R.L., Kroenke, K., & Williams, J.B. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA, 282*(18), 1737–1744. https://doi.org/10.1001/jama.282.18.1737

Spitzer, R.L., Kroenke, K., Williams, J.B., & Lowe, 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

Stander, M.P., Bergh, M., Miller-Janson, H.E., De Beer, J.C., & Korb, F.A. (2016). Depression in the South African workplace. *South African Journal of Psychiatry, 22*(1), Article a814. https://doi.org/10.4102/sajpsychiatry.v22i1.814

Stein, D.J., Seedat, S., Herman, A., Moomal, H., Heeringa, S.G., Kessler, R.C., & Williams, D.R. (2008). Lifetime prevalence of psychiatric disorders in South Africa. *British Journal of Psychiatry,* 192(2), 112–117. https://doi.org/10.1192/bjp.bp.106.029280

Swartz, L. (1998). *Culture and mental health: A southern African view*. Cape Town: Oxford University Press.

Tomlinson, M., Grimsrud, A.T., Stein, D.J., Williams, D.R., & Myer, L. (2009). The epidemiology of major depression in South Africa: Results from the South African Stress and Health study. *South African Medical Journal, 99*(5 Pt 2), 367–373.

Van Wijk, C.H., Cronje, F.J., & Meintjes, W.A.J. (2020). Mental wellbeing monitoring in a sample of Emergency Medical Service personnel. *Occupational Diseases and Environmental Medicine, 8*(1), 26–33. https://doi.org/10.4236/odem.2020.81002

Van Wijk, C.H., Martin, J.H., & Meintjes, W.A.J. (2021). Burden of common mental disorders in South African workplace settings. *Occupational Health Southern Africa.* Manuscript submitted for publication.

Vissoci, J., Hertz, J., El-Gabri, D., Andrade Do Nascimento, J.R., Pestillo De Oliveira, L., Mmbaga, B.T. …Staton, C.A. (2018). Cross-cultural adaptation and psychometric properties of the AUDIT and CAGE questionnaires in Tanzanian Swahili for a traumatic brain injury population. *Alcohol and Alcoholism, 53*(1), 112–120. https://doi.org/10.1093/alcalc/agx058

Volker, D., Zijlstra-Vlasveld, M.C., Brouwers, E.P., Homans, W.A., Emons, W.H., & Van der Feltz-Cornelis, C.M. (2016). Validation of the Patient Health Questionnaire-9 for major depressive disorder in the occupational health setting. *Journal of Occupational Rehabilitation, 26*(2), 237–244. https://doi.org/10.1007/s10926-015-9607-0

Wang, J., Kelly, B.C., Booth, B.M., Falck, R.S., Leukefeld, C., & Carlson, R.G. (2010). Examining factorial structure and measurement invariance of the Brief Symptom Inventory (BSI)-18 among drug users. *Addiction Behaviour, 35*, 23–29. Retrieved from https://psycnet.apa.org/doi/10.1016/j.addbeh.2009.08.003

Wang, J., & Wang, X. (2020). *Structural equation modeling: Applications using Mplus* (2nd ed.). New York: John Wiley & Sons.

Ward, C.L., Lombard, C.J., & Gwebushe, N. (2006). Critical incident exposure in South African emergency services personnel: Prevalence and associated mental health issues. *Emergency Medicine Journal, 23*(3), 226–231. https://doi.org/10.1136/emj.2005.025908

Weobong, B., Akpalu, B., Doku, V., Owusu-Agyei, S., Hurt, L., Kirkwood, B., & Prince, M. (2009). The comparative validity of screening scales for postnatal common mental disorder in Kintampo, Ghana. *Journal of Affective Disorders, 113*(1–2), 109–117. https://doi.org/10.1016/j.jad.2008.05.009

Williams, N. (2014). The CAGE questionnaire. *Occupational Medicine, 64*(6), 473–474. https://doi.org/10.1093/occmed/kqu058

Wilson, A., Wissing, M.P., & Schutte, L. (2018). Validation of the Stress Overload Scale and Stress Overload Scale–short form among a Setswana-speaking community in South Africa. *South African Journal of Psychology, 48*(1), 21–31. https://doi.org/10.1177%2F0081246317705241

Zhong, Q.Y., Gelaye, B., Zaslavsky, A.M., Fann, J.R., Rondon, M.B., Sánchez, S.E., & Williams, M.A. (2015). Diagnostic validity of the Generalized Anxiety Disorder-7 (GAD-7) among pregnant women. *PLoS One, 10*(4), e0125096. https://doi.org/10.1371/journal.pone.0125096

Zungu, L.I. (2013). Prevalence of post-traumatic stress disorder in the South African mining industry and outcomes of liability claims submitted to Rand Mutual Assurance Company. *Occupational Health Southern Africa, 19*(2), 22–26. https://doi.org/10.1186%2F1471-244X-13-182