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

Differential item functioning of the CESDR-R and GAD-7 in African and white working adults

Authors:
Carolina Henn[1]
Brandon Morgan[1]

Affiliations:
[1]Department of Industrial Psychology and People Management, College of Business and Economics, Faculty of Management, University of Johannesburg, Auckland Park, South Africa

Corresponding author:
Brandon Morgan,
bmorgan@uj.ac.za

Dates:
Received: 20 Mar. 2019
Accepted: 24 July 2019
Published: 10 Oct. 2019

How to cite this article:
Henn, C., & Morgan, B. (2019). Differential item functioning of the CESDR-R and GAD-7 in African and white working adults. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde, 45*(0), a1663. https://doi.org/10.4102/sajip.v45i0.1663

Orientation: Depression and anxiety can have undesirable consequences for employees and their employers. It is therefore important that employers pay attention to the existence and extent of depression and anxiety. However, measuring these constructs requires unbiased, reliable and valid instruments.

Research purpose: To facilitate unbiased measurement of depression and anxiety, we investigated differential item functioning of the Centre for Epidemiologic Studies Depression Scale-Revised (CESD-R) and Generalised Anxiety Disorder Scale 7 (GAD-7) in a sample of non-clinical African and white working adults.

Motivation for the study: Biased measurement instruments can lead to serious problems when comparing scores between groups, using raw score cut-offs, or creating norm scores. Practitioners are legally and ethically required to ensure that any instrument used is unbiased.

Research approach/design and method: A cross-sectional survey design was used. The CESD-R and GAD-7 were administered to working adults. A final sample of 551 CESD-R responses and 529 GAD-7 responses were included in the analyses. Ordinal logistic regression was performed to investigate differential item functioning.

Main findings: Both CESD-R and GAD-7 showed some evidence of differential item functioning although it was mostly small in magnitude. Item bias had some minor non-negligible impact on aggregated observed scores within specific ranges of the underlying traits.

Practical/managerial implications: Both CESD-R and GAD-7 show promise as instruments that can be utilised to explore the experience of anxiety and depression in African and white employees.

Contribution/value-add: This study is a promising first step towards the measurement fairness of the CESD-R and GAD-7 in the South African context.

Keywords: Depression; Anxiety; GAD-7; CESD-R; Differential Item Functioning.

Introduction

Orientation

The World Health Organization (2014) defines mental health as:

> [*A*] state of well-being in which every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community. (para. 1)

It is important that mental illness is identified and effectively treated to promote and maintain mental health. Statistics with regard to mental illness are sobering. The World Health Assembly (2012) reports that 76% – 85% of persons with mental illnesses in low- to middle-income countries do not receive treatment. Poor mental health has social and economic impacts, directly affecting an individual’s ability to work and earn an income, and indirectly affecting the economy at a national level (World Health Assembly, 2012).

Depression and anxiety disorders are considered to be among the top 10 causes of workplace disability globally (Harnois & Gabriel, 2002). Both mental health conditions pose a real threat to employee well-being and organisational effectiveness (Bender & Farvolden, 2008; Evans-Lacko et al., 2016). A workplace is 'an ideal setting for depression and anxiety interventions' (Mykletun & Harvey, 2012, p. 868) because adults spend much of their time in the workplace (Tan et al., 2014). Indeed, research supports the effectiveness of workplace interventions for depression and anxiety (Joyce, Modini, Christensen, & Mykletun, 2016; Tan et al., 2014). Despite these findings, anxiety and depression in the workplace appear to have received little research attention, particularly in South Africa, with limited research available on their workplace impact. Therefore, it is important that more research is conducted on these topics. However, to be able to do this, it is necessary to measure depression and anxiety in the workplace with appropriate, psychometrically sound and unbiased measuring instruments. The Centre for Epidemiologic Studies Depression Scale (CESD) (Radloff, 1977) and its revised version (CESD-R) (Eaton, Smith, Ybarra, Muntaner, & Tien, 2004) and the Generalised Anxiety Disorder Scale 7 (GAD-7) (Spitzer, Kroenke, Williams, & Löwe, 2006) are often used to measure depression and anxiety in clinical and non-clinical samples. These two instruments are particularly useful as they measure the symptoms most commonly associated with depression and anxiety.

Research purpose and objectives

Unfortunately there is limited evidence available on the reliability and validity of these two instruments when used in South African workplaces. It is particularly important that measurement bias, also referred to as Differential Item Functioning (DIF), in these two instruments as one form of validity is investigated. Differential item functioning means that respondents from different groups[1] who have the same relative standing on some latent trait have different response probabilities on one or more items of an instrument that measures the latent traits (Chalmers, Counsell, & Flora, 2016; Zumbo, 1999). These differences (after respondents are matched on the latent traits) usually reflect construct irrelevant factors that can confound observed scores (De Sa-Junior et al., 2019). This is problematic because decisions are made on observed scores with the assumption that these scores are uncontaminated by unwanted sources of variance (e.g. Gamerman, Gonçalves, & Soares, 2018; Steyn & De Bruin, 2018).

Differential item functioning often leads to item bias (Sireci, 2011) and failure to account for this bias can lead to biased or even incorrect decisions when item scores are translated into aggregated scale scores. Investigating DIF is therefore both a legal and an ethical imperative (Jodoin & Gierl, 2001). In South Africa, the *Health Professions Act* of 1974 (Republic of South Africa, 1974) and the *Employment Equity Act* (EEA) of 1998 (Republic of South Africa, 1998) are clear on the legal and ethical imperatives of using unbiased instruments, with the EEA indicating that no instruments (or scores obtained from these instruments) should be used in the workplace if they are unreliable, have limited validity evidence, and/or are biased against any person. It is the responsibility of practitioners to 'contribute to specific empirical studies related to the psychometric properties of the tests they use' (Health Professions Council of South Africa Form 208, 2006, p. 1) and to ensure that there is '[*e*]vidence relating to … DIF … [*w*]*hen tests are to be used with individuals from different groups*' (italics in original, International Test Commission, 2013, p. 17).

Applied to the CESD-R and GAD-7 item bias (i.e. DIF) means that observed scores between groups cannot be directly compared because these scores consist of construct relevant and construct irrelevant sources of variance (De Sá Junior et al., 2019; Van De Vijver & Leung, 2011; Zumbo, 1999). It also holds implications for cut-off scores used to indicate depression and anxiety (Carleton et al., 2013; Spitzer et al., 2006) and norm scores created for these instruments (Carleton et al., 2013). These two instruments must therefore be free of DIF if they are used to measure depression and anxiety in South African workplaces. Against this background and as a starting point for future research, this study sets out to investigate DIF in the CSED-R and GAD-7 items with a sample of non-clinical African and white working adults. Unfortunately, we could not include mixed-race, Indian and/or Asian participants in our study because there was insufficient data available to allow for meaningful DIF analysis. The results of this study hold important implications for the psychometric properties of the CESD-R and GAD-7 and their use in South African workplaces. In the following sections, we provide a brief overview of depression and anxiety at work. We then discuss the CESD-R and GAD-7 and distinguish between item bias and DIF.

Literature review

Depression and anxiety at work

Depression is a mood disorder that includes symptoms such as feelings of worthlessness, depressed mood, decreased or increased appetite, insomnia or hypersomnia, poor concentration, anhedonia, loss of energy and loss of interest (American Psychiatric Association [APA], 2013). Generalised anxiety is characterised by symptoms such as feelings of worry and apprehension, feeling constantly on edge, feelings of impending doom and physical sensations such as heart palpitations, sweating hands and nausea (refer to the Diagnostic and Statistical Manual for Mental Disorders [DSM-V], APA, 2013). In South Africa, the lifetime prevalence of depression and anxiety is approximately 9.8% and 5.8%, respectively (Herman et al., 2009).

Given the typical symptoms of anxiety and depression, and the relatively high prevalence thereof in South Africa, it is to be expected that both these mental disorders would have a detrimental effect in the workplace. These effects include an increased risk of workplace accidents, a deterioration in work

[1] Group is defined broadly and could include, for example, 'groups differing in terms of gender, cultural background, education, ethnic origin, or age' (International Test Commission, 2013, p. 17). Zumbo (1999, p. 13) indicates that 'standard comparisons are based on gender, race, sub-culture, or language'.

performance and increased absenteeism (Haslam, Atkinson, Brown, & Haslam, 2005). Mall et al. (2015), for example, reported that depression and anxiety, respectively, caused 27.2 and 28.2 days out-of-role annually. Organisations are impacted negatively by factors such as decreased productivity and high staff turnover (Haslam et al., 2005). Workplaces can also worsen symptoms of depression because of factors such as high workload and stigmatised attitudes towards mental illness (Haslam et al., 2005). Given the ubiquitousness of depression and anxiety and their effects at work, it is crucial that organisations give high priority to these mental disorders. According to Mall et al. (2015), however, this is not yet the case. Organisations need to investigate depression and anxiety in their organisations, and to be able to do so it is important that psychometrically sound measuring instruments are used. Moreover, in South Africa with its multicultural context, such measuring instruments should be unbiased (i.e. fair) so that scores do not disadvantage any particular group.

The Centre for Epidemiologic Studies Depression Scale-Revised

Radloff (1977) developed the Centre for Epidemiologic Studies Depression Scale (CES-D), a self-report instrument, to measure depression in the general population. The CES-D has been used extensively in research. In fact, in their search of the Psych Articles Database (PAD), Van Dam and Earleywine (2010) found that the Beck Depression Inventory (BDI) and the CES-D are the two most commonly used depression scales. It has since been revised (Eaton et al., 2004) to reflect the diagnostic criteria for depression described in the DSM-IV-TR (APA, 2000). In the most recent update of the DSM, the DSM-V (APA, 2013), the diagnostic criteria for depression remain the same as in the DSM-IV-TR. Eaton et al. (2004) reported a correlation of 0.93 between the CESD and the CESD-R scores, meaning that these instruments are mostly interchangeable. Although the revision to the CES-D has been done in 2004, to our knowledge little research has been published on the CESD-R.

The CESD-R (Eaton et al., 2004) consists of 20 items. For each item, participants must indicate how often during the past 2 weeks they have felt or behaved in a particular way. With regard to factor structure of the CESD-R, Eaton et al. (2004) and Van Dam and Earleywine (2010) found support for a unidimensional model. Walsh (2014), on the other hand, found good model fit for a two-factor model with 16 items instead of the original 20 items. Unfortunately, Walsh (2014) did not report on the full-factor pattern matrix or factor correlation matrix, making it difficult to determine how much shared variance exists in these two factors. Van Dam and Earleywine (2010) and Walsh (2014) established discriminant and convergent validity for the CESD-R. Internal consistency coefficients typically are all above 0.90 (Eaton et al., 2004; Van Dam & Earleywine, 2010; Walsh, 2014). In South Africa, the CESD-R was validated for use among a South African non-clinical sample of working adults (Michas & Henn, 2019). Michas and Henn (2019) found good model fit for a one-factor model and convergent and discriminant validity was also established. Reliability of the scale score was also supported with an internal consistency coefficient of 0.95.

The Generalised Anxiety Disorder Scale

The GAD-7 (Spitzer et al., 2006, p. 1092) was developed to 'identify probable cases of GAD and to assess symptom severity'. Items were developed based on diagnostic criteria in the DSM-IV (APA, 2005) as well as the investigation of existing anxiety scales. Although originally intended for clinical and primary care settings, the instrument gained popularity as a psychometrically sound instrument to measure and detect generalised anxiety in the general population. The GAD-7 is a self-report questionnaire with seven items. Participants on this scale indicate how often the item stems occurred over the last 2 weeks (Spitzer et al., 2006).

Many validation studies of the GAD-7 have been undertaken in a range of populations, such as pregnant women (Zhong et al., 2015), the general population (Löwe et al., 2008), infertile men and women (Omani-Samani, Maroufizadeh, Ghaberi, & Navid, 2018), outpatients with diagnosed anxiety and mood disorders (Rutter & Brown, 2017), the psychiatric population (Beard & Björgvinsson, 2014), employees (Henn & Bezuidenhout, 2019), adolescents (Tiirikainen, Haravuori, Ranta, Kaltiala-Heino, & Marttunen, 2019) and primary care patients (Jordan, Shedden-Mora, & Löwe, 2017). The scale has also been translated into several other languages, including Spanish (García-Campayo et al., 2010), German (Löwe et al., 2008), French (Barthel, Barkmann, Ehrhardt, Bindt, & International CDS Study Group, 2014), Twi (Barthel et al., 2014) and Dutch (Donker, Van Straten, Marks, & Cuijpers, 2011).

With regard to factor structure, most studies found a unidimensional model to have the best fit (e.g. García-Campayo et al., 2010; Hinz et al., 2017; Jordan et al., 2017; Löwe et al., 2008; Omani-Samani, et al., 2018; Sousa et al., 2015; Tiirikainen et al. 2019). Barthel et al. (2014) found that although a unidimensional model performed the best, the explained variance was rather low. A non-negligible structure was found in the standardised residuals after fitting the Rasch model, suggesting that the scale might not necessarily be unidimensional in their sample. In a cross-cultural study, Parkerson, Thibodeau, Brandt, Zvolensky and Asmundson (2015) found a unidimensional model for white participants only. They also found some evidence for DIF in items GAD-7 1, GAD-7 5 and GAD-7 6, with black or African American participants generally showing lower expected scores after matching on the latent traits compared to Hispanic and white participants. The DIF tended to be most pronounced at the upper end of the latent trait.

Henn and Bezuidenhout (2019) validated the GAD-7 for use in a non-clinical sample of employees in South Africa and found a good model fit for a one-factor model and also reported evidence of discriminant validity and some convergent validity. Reported alpha coefficients are generally higher, for example, 0.92 (Spitzer et al., 2006), 0.93 (Garcia-Campayo et al., 2010), 0.88 (Sousa et al., 2015), 0.89 (Zhong et al., 2015), 0.91 (Tirrikainen et al., 2018) and, in South Africa it is 0.92 (Henn & Bezuidenhout, 2019). Barthel et al. (2014), however, reported lower alpha coefficients of 0.69 for French-speaking persons in Côte d'Ivoire, and 0.67 for Twi-speaking persons in Ghana.

Differential item functioning

We previously defined DIF as different response probabilities to an item across groups when participants in these groups are matched on the latent traits (Zumbo, 1999). Van De Vijver and Leung (2011) state that:

> [*A*]n item is biased [*i.e., has DIF*] if respondents with the same standing on the underlying construct [*i.e., the latent trait*] … do not have the same mean [*or expected*] score on the item because of different cultural origins. (p. 25)

These two definitions are equivalent. Differential item functioning and item bias are often used interchangeably. However, as Sireci (2011) points out, an item should be considered biased when it shows non-negligible DIF and when this DIF can be ascribed to construct irrelevant factors. In other words, this DIF is because of 'some characteristic of the test item or testing situation that is not relevant to the test purpose' (Zumbo, 1999, p. 12). In the DIF literature, a distinction is often made between uniform and non-uniform DIF. Uniform DIF implies that differences in the probability of item endorsement across groups conditional on latent trait estimates are constant across the whole distribution. Non-uniform DIF occurs when these differences change (i.e. interact) at different locations in this distribution (Berger & Tutz, 2015; De Beer, 2004; Gamerman et al., 2018; Jodoin & Gierl, 2001). In the next section, we detail the method used in this study.

Method

Research approach

Data were collected from four studies under the supervision of the first author (Claassens, 2018; Michas, 2018; Sekatane, 2018; Tsebe, 2018). These studies used a quantitative research approach and a cross-sectional survey research design. A cross-sectional design allowed the researchers to investigate and interpret results from participants at the same point in time (Gravetter & Forzano, 2015).

Research participants

Non-probability convenience and snowball sampling methods were used in the aforementioned studies to obtain participants. Convenience sampling is employed when any person meeting the inclusion criteria can be invited to participate in the study. It was extended in this study to snowball sampling as participants referred other potential participants to the researchers (Gravetter & Forzano, 2015). These sampling methods were utilised because the working population in South Africa is large and it was therefore not possible to obtain a random sample. It also aided in obtaining a heterogeneous sample from a wide variety of industries. Working adults who were 18 years old or above, able to read and write in English and employed for at least 1 year were invited to participate in the study. In total, 687 responses to the GAD-7 and CESD-R were obtained. Because of data cleaning and merging of different data sets there were some differences in the final sample groups used in our analyses. We therefore provide a description of the sample used for the CESD-R analysis and the sample used for the GAD-7 analysis.

After cleaning the data we had 551 CESD-R scores of African (n = 307) and white (n = 244) participants. As mentioned earlier, other race groups were not included because there was insufficient data available to meaningfully investigate DIF. The mean age of the participants was 36.15 years (median = 33, Standard Deviation [SD] = 11.26). For the African participants, the mean age was 34.15 years (median = 32, SD = 8.97) and for the white participants the mean age was 38.54 years (median = 36, SD = 13.12). There were approximately twice as many women (n = 346, 64.19%) as men (n = 193, 35.81%) in the sample for both the African (women: n = 186, 63.05%; men: n = 109, 36.95%) and white (women: n = 160, 65.67%; men: n = 84, 34.43%) sample groups. Most of the participants indicated that their home language was Afrikaans (n = 149, 27.04%) or Sepedi (n = 116, 21.05%). The mean years of employment was 8.11 (median = 5, SD = 8.23) for the African participants and 8.83 (median = 6, SD = 9.01) for the white participants.

We had 529 GAD-7 scores of African (n = 304) and white (n = 225) participants. The mean age of the participants was 38.83 years (median = 35, SD = 13.26). For the African participants, the mean age was 34.18 years (median = 34.18, SD = 9.38) and for the white participants the mean age was 38.83 years (median = 36, SD = 13.26). There were approximately twice as many women (n = 339, 65.44%) as men (n = 179, 34.56%) in the sample for both the African (women: n = 185, 63.13%; men: n = 108, 36.86%) and white (women: n = 154, 68.44%; men: n = 71, 31.56%) sample groups. Most of the participants stated that their home language was Afrikaans (n = 140, 26.47%) or Sepedi (n = 111, 20.98%). The mean years of employment was 8.40 (median = 5, SD = 8.37) for the African participants and 8.77 (median = 5, SD = 9.03) for the white participants.

Measuring instruments

The CESD-R (Eaton et al., 2004) and GAD-7 (Spitzer et al., 2006) scales were used in this study. The CESD-R consists of 20 items and participants have to indicate how often they have behaved in a particular way or experienced a particular feeling in the past 1 week or so. Responses are based on a five-point scale ranging from *not at all or less than 1 day* to *nearly every day for 2 weeks*. The GAD-7 has seven items that are scored on a four-point scale ranging from *not at all* to *nearly every day*. Participants have to indicate how often over the last 2 weeks they have been bothered by particular problems. On both scales, higher scores indicate higher levels of depression and anxiety, respectively. As these instruments were presented in detail in the literature review section, no further information will be provided here.

Research procedure and ethical considerations

The data used in this study were collected during 2016 and 2017 as part of a larger project on mental health in the workplace. Data were collected online and in person using hard copy questionnaires. Participants were informed of the purpose of the study, that the participation was voluntary and that they were free to withdraw from the study at any point in time without any adverse consequences. No identifying information was obtained, ensuring anonymity and confidentiality. The participants gave consent for their responses to be used in future studies. The contact details of the first author, who is a registered counselling psychologist, was provided to the participants in the event that they required any psychological assistance.

Statistical analysis

Differential item functioning was investigated using ordinal logistic regression as implemented in the *lordif* package (Choi, Gibbons, & Crane, 2011, version 0.3-3) in *R* (R Core Team, 2018, version 3.4.1, Vienna, Austria). This approach uses a series of nested logistic regression models for each item to investigate DIF, where the items and their associated ordered response categories (probability of endorsement of a response category) are the outcome variables. Model 1 uses trait level as a predictor. Trait level in this context represents the latent variable score estimates (thetas) for the CESD-R and GAD-7. Model 2 uses trait level and group membership as predictors and Model 3 uses trait level, group membership and their interaction as predictors (Choi et al., 2011; Crane, Gibbons, Jolley, & Van Belle, 2006). The graded response model (Samejima, 1969) was fit to item responses and trait-level estimates obtained using an iterative purification procedure (see Crane et al., 2006 for an overview of this technique). The iterative purification procedure approach can help reduce the detection of artificial DIF (Hagquist & Andrich, 2017).

We used the likelihood ratio test for models 1 and 2 to investigate uniform DIF, models 2 and 3 to investigate non-uniform DIF and models 1 and 3 to investigate total DIF. Statistical significance for each likelihood ratio test was set to $p < 0.01$ instead of the usual $p < 0.05$. We did this to account for multiple comparisons while still preserving power to detect potential DIF (Hope, Adamson, McManus, Chis, & Elder, 2018). To assist in detecting uniform DIF, we also investigated the proportional change in the beta coefficient of trait level between models 1 and 2 for each item (Choi et al., 2011). Following Crane et al. (2007), we used a proportional change of 0.05 (5%) as our cut-off value and compared these results to the aforementioned likelihood ratio tests. Lastly, we compared the difference in Nagelkerke's pseudo R^2 across the three models as an approximate measure of the magnitude of the DIF for each item. Jodoin and Gierl's (2001) criteria were used, where $\Delta R^2 < 0.035$ indicates a negligible DIF, ΔR^2 between 0.035 and 0.070 indicates a moderate DIF and $\Delta R^2 > 0.070$ indicates a large DIF. For each analysis the African participants were the reference group because the African group had a larger sample size and the white participants were the focus group.

Results

Descriptive statistics and reliability coefficients

Descriptive statistics and reliability coefficients for the GAD-7 and CESD-R scale scores for each group are presented in Table 1. Satisfactory reliability coefficients were found for both scale scores. We applied Revelle's coefficient β (Revelle, 1979) to investigate unidimensionality of each scale. Previous studies (Henn & Bezuidenhout, 2019; Michas & Henn, 2019) using parts of these data have already established unidimensionality of the CESD-R and GAD-7 scales. Coefficient β was therefore used as an additional descriptive measure of unidimensionality of the item scores in our analysis. In brief, coefficient β indicates the proportion of variance in item scores that can be attributed to a common factor (Revelle & Zinbarg, 2008). It should at a minimum be 0.50 (Revelle, 1979). Unidimensionality was supported for both the CESD-R and GAD-7 scales, with β coefficients all > 0.50. Coefficient β was somewhat smaller for the African CESD-R responses, suggesting that there might be some deviation from unidimensionality. However, it was not considered to be of practical concern in this study because it still suggested unidimensionality.

Differential item functioning Centre for Epidemiologic Studies Depression Scale-Revised

Table 2 presents the DIF results for the CESD-R. Data in the table show that six (30%) of the 20 CESD-R items had statistically significant total DIF (i.e. χ^2 Model 1 and Model 3). Statistical significance at $p < 0.05$ was retained for three items after applying Holm–Bonferroni corrections to the p values. Five items showed uniform DIF and one item showed non-uniform DIF. However, the proportional changes in beta

TABLE 1: Descriptive statistics and reliability coefficients for the Centre for Epidemiological Depression Scale Revised and Generalised Anxiety Disorder 7 scale scores across African and white participants.

Participants	Mean	Mdn.	SD	Skew.	Kurt.	SE	α	ω	β
CESD-R African	15.37	11.00	14.33	1.27	1.17	0.82	0.94	0.92, 0.95	0.69
CESD-R White	14.04	10.00	14.29	1.87	3.90	0.92	0.95	0.93, 0.96	0.80
GAD-7 African	5.91	4.00	5.34	0.89	-0.09	0.31	0.91	0.89, 0.93	0.86
GAD-7 White	6.28	5.00	5.31	0.89	-0.06	0.35	0.93	0.92, 0.95	0.85

Note: 95% bootstrapped confidence intervals for coefficient α and ω in parentheses. CESD-R African n = 307, white n = 244. GAD-7 African n = 304, white n = 225.

CESD-R, Centre for Epidemiological Depression Scale Revised; GAD-7, Generalised Anxiety Disorder 7-Item Scale; SD, standard deviation; SE, standard error; Mdn., median; Skew., skewness; Kurt., kurtosis; β, Revelle's coefficient beta.

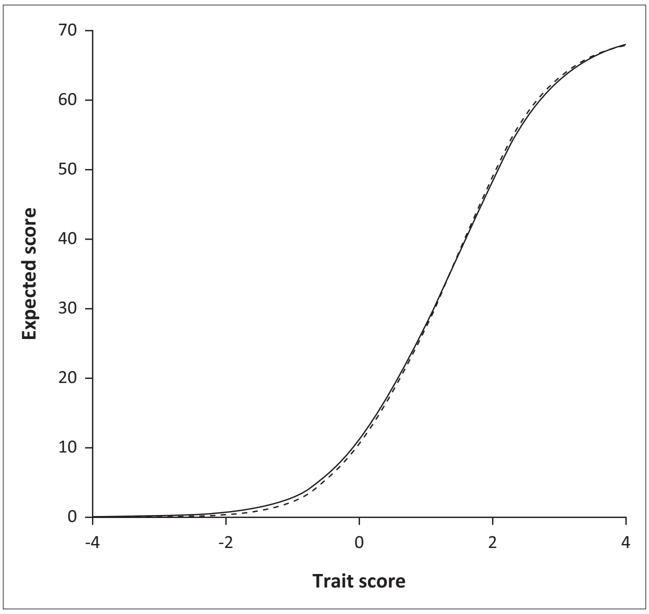
TABLE 2: Differential item functioning for the Centre for Epidemiological Depression Scale Revised items.

Item	χ^2_{M12}	χ^2_{M23}	χ^2_{M13}	p adj.	ΔR^2_{M12}	ΔR^2_{M23}	ΔR^2_{M13}	$\Delta\beta$
CESD-R 1	**0.001**	0.656	**0.003**	0.054	0.018	0.000	0.018	0.002
CESD-R 2	0.870	**0.000**	**0.002**	**0.034**	0.000	0.011	0.011	0.001
CESD-R 3	0.230	0.048	0.069	0.898	0.001	0.003	0.004	0.003
CESD-R 4	0.645	0.395	0.626	1.000	0.000	0.001	0.001	0.002
CESD-R 5	0.120	0.121	0.090	1.000	0.003	0.003	0.005	0.006
CESD-R 6	**0.009**	0.033	**0.003**	0.054	0.006	0.004	0.009	0.004
CESD-R 7	**0.000**	0.581	**0.000**	**0.000**	0.016	0.000	0.016	0.045
CESD-R 8	0.029	0.045	0.012	0.174	0.005	0.004	0.009	0.002
CESD-R 9	0.934	0.464	0.762	1.000	0.000	0.001	0.001	0.000
CESD-R 10	0.400	0.593	0.608	1.000	0.001	0.000	0.001	0.001
CESD-R 11	**0.000**	0.583	**0.000**	**0.000**	**0.041**	0.001	**0.042**	0.018
CESD-R 12	0.763	0.894	0.947	1.000	0.000	0.000	0.000	0.001
CESD-R 13	0.449	0.853	0.738	1.000	0.001	0.000	0.001	0.005
CESD-R 14	0.881	0.071	0.194	1.000	0.000	0.009	0.009	0.001
CESD-R 15	0.186	0.373	0.281	1.000	0.006	0.003	0.009	0.000
CESD-R 16	0.725	0.604	0.821	1.000	0.000	0.000	0.000	0.001
CESD-R 17	**0.003**	0.373	**0.008**	0.122	0.014	0.001	0.015	0.029
CESD-R 18	0.043	0.858	0.126	1.000	0.009	0.000	0.010	0.006
CESD-R 19	0.597	0.180	0.354	1.000	0.000	0.002	0.003	0.002
CESD-R 20	0.520	0.410	0.579	1.000	0.000	0.001	0.001	0.003

Note: Values reported under χ^2 columns are the p-values for the likelihood ratio tests. Statistically significant p-values, $\Delta R^2 > 0.035$ and $\Delta\beta > 0.05$ in bold.
M12, models 1 and 2; M23, models 2 and 3; M13, models 1 and 3; p adj., Holm–Bonferroni adjusted p-values for χ^2_{M13}; $\Delta\beta$, proportional changes in β.

coefficients were < 0.050 for all five of these items showing uniform DIF and only one item had a $\Delta R^2 > 0.035$. The mean difference in initial (not accounting for DIF) and purified (accounting for DIF) theta estimates across both groups was 0.000 (median = -0.000, SD = 0.026, range = -0.098 to 0.101, interquartile range = -0.016 to 0.020). Items CESD-R 7 and CESD-R 11 appeared to be especially problematic, with the item characteristic curves showing that white participants were more likely to endorse item CSED-R 7 and less likely to endorse item CSED-R 11 across the whole latent distribution.

Figure 1 presents the test characteristic curve using group-specific item parameters. The figure shows that there were minor overall differences in expected scores, suggesting that expected score differences in opposite directions at the item level (i.e., over- and under-estimation) cancelled each other out. White participants had slightly lower expected scores at the lower end and slightly higher expected scores at the higher end of the theta distribution. As a whole, however, these results suggest that DIF had a negligible overall impact on expected scores.[2]



Solid black line, African participants; dashed black line, white participants.

FIGURE 1: Test characteristic curve for all Centre for Epidemiologic Studies Depression Scale-Revised items.

Differential item functioning Generalised Anxiety Disorder Scale-7

Table 3 presents the DIF results for the GAD-7. Data in the table show that two (29%) of the seven GAD-7 items had statistically significant total DIF (i.e. χ^2 Model 1 and Model 3). Statistical significance for these two items at $p < 0.05$ was retained after applying Holm–Bonferroni corrections to the p values. Both of the identified items showed uniform DIF. However, the proportional changes in beta coefficients were < 0.050 and the ΔR^2 were < 0.035. Item GAD-7 6 did not reach statistical significance at $p = 0.01$. It, however, showed potential uniform DIF with a p value of 0.012 (and 0.065 after applying the Holm–Bonferroni corrections).

The mean difference in initial (not accounting for DIF) and purified (accounting for DIF) theta estimates was 0.000 (median = -0.000, SD = 0.037, range = -0.152 to 0.157, interquartile range = -0.021, 0.013). Item characteristic curves showed that white participants were more likely to endorse

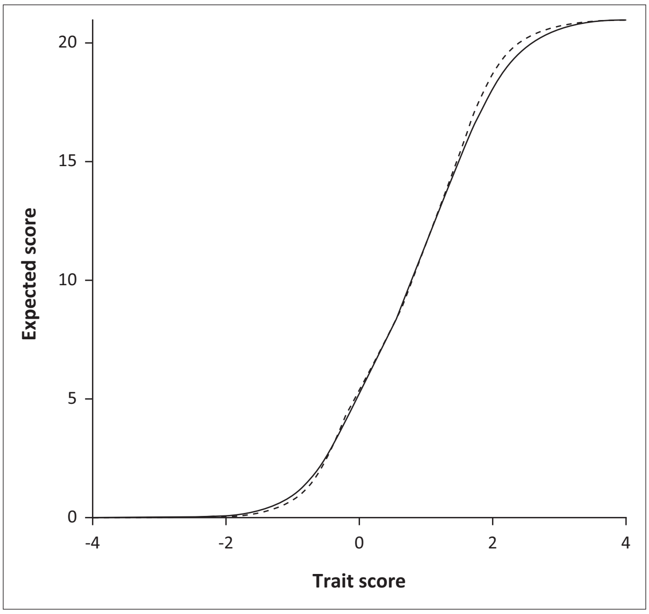
[2] To support our findings we also investigated differential test functioning using the procedures described by Chalmers et al. (2016) as implemented in the *mirt* (Chalmers, 2012) package version 1.30. The results showed that there was approximately 0.50% (0.32%, 1.10%) average absolute difference in test response curves (integrated over a theta range of -4.00 to 4.00) and approximately 0.03 (-0.46, 0.53, p = 0.91) raw score bias on average. Both of these scores indicate negligible overall impact of DIF at the aggregated score level. Plotting the signed differential test functioning suggested that there was minor non-negligible DIF in the -3.00 to 0.40 theta range, with a maximum of 0.70 raw score bias in this range. These results can be obtained from the second author.

TABLE 3: Differential item functioning for the Generalised Anxiety Disorder 7 items.

Item	χ^2_{M12}	χ^2_{M23}	χ^2_{M13}	*p* adj.	ΔR^2_{M12}	ΔR^2_{M23}	ΔR^2_{M13}	$\Delta\beta$
GAD-7 1	**0.002**	0.626	**0.007**	**0.043**	0.007	0.000	0.007	0.006
GAD-7 2	0.976	0.183	0.412	0.825	0.000	0.001	0.001	0.000
GAD-7 3	**0.007**	0.093	**0.006**	**0.043**	0.004	0.002	0.006	0.018
GAD-7 4	0.804	0.582	0.833	0.833	0.000	0.000	0.000	0.000
GAD-7 5	0.069	0.240	0.096	0.287	0.003	0.001	0.004	0.008
GAD-7 6	0.012	0.298	0.025	0.125	0.006	0.001	0.007	0.004
GAD-7 7	0.117	0.066	0.054	0.215	0.002	0.003	0.005	0.007

Note: Values reported under χ^2 columns are the *p*-values for the likelihood ratio tests. Statistically significant *p*-values, $\Delta R^2 > 0.035$ and $\Delta\beta > 0.05$ in bold.
M12, models 1 and 2; M23, models 2 and 3; M13, models 1 and 3; *p* adj., Holm–Bonferroni adjusted *p*-values for χ^2_{M13}; $\Delta\beta$, proportional changes in β.

item GAD-7 1 and less likely to endorse item GAD-7 3 for most of the underlying traits. Item GAD-7 1 appeared to be especially problematic at the upper end of the theta distribution. Figure 2 provides the test characteristic curve using group specific item parameters. An inspection of the figure shows that there were minor differences in expected scores at the lower end of the theta distribution. The differences in expected scores were more pronounced at the upper end of the theta distribution, with white participants having larger expected scores. As a whole, these results suggest that DIF had a minor overall impact on expected scores at the lower end and middle of the theta distribution. However, DIF appeared to have a larger effect at the upper end of the theta distribution.[3]



Solid black line, African participants; dashed black line, white participants.

FIGURE 2: Test characteristic curve for all Generalised Anxiety Disorder Scale 7 items.

Ethical considerations

For all of the studies, ethical clearance was obtained from the then Faculty of Management Research Ethics Committee at the University of Johannesburg. The ethical clearance codes were FOM-2016IPPM029, FOM-2016IPPM032, IPPM-2017-103 (M) and IPPM-2017-104 (M).

Discussion

This study set out to investigate DIF of the CESD-R and GAD-7 in a sample of non-clinical African and white working adults as a starting point for future research on the validity of these two instruments in the South African context. The objective of the study was achieved. The results showed that six of the 20 CESD-R items and two of the seven GAD-7 items showed statistically significant DIF. However, the DIF magnitudes were quite small with respect to the change in R^2, suggesting that DIF had little impact at the item level. Parkerson et al. (2015) also found DIF for item GAD-7 1 in the United States and in the same direction as our results. This item therefore certainly requires some attention. The test characteristic curve for the CESD-R showed that DIF at the item level did not translate to meaningful DIF in aggregated scale scores. The test characteristic curve for the GAD-7 showed similar results although there was some evidence for minor non-negligible DIF in the upper end of the theta distribution. Few participants scored high on the GAD-7 despite reducing precision of estimates in this range. Overall these differences are probably too small to make any real difference when using the CESD-R and GAD-7 in research settings for measuring participants across the whole latent distribution. The same might not be true when using these instruments in practice.

It is noteworthy that DIF appeared to have a non-negligible impact at the upper end of the GAD-7 trait scores because Parkerson et al. (2015) obtained similar results. Although the expected difference in scores was not large, these results suggest that there might be some bias for African and white participants who score high on anxiety when using this instrument. This minor non-negligible DIF should not therefore be ignored when using the GAD-7 for screening purposes, especially if decisions are based on the cut-off scores used in the literature (e.g. Carleton et al., 2013). As indicated by Parkerson et al. (2015), practitioners should be aware of this potential bias because it can lead to over- or underestimation of trait-level scores. The minor non-negligible bias in the CESD-R scores should not also be routinely ignored when using scores for screening purposes although the overall impact on aggregated scale scores is less pronounced than the GAD-7.

[3] To support our findings we also investigated differential test functioning using the procedures described by Chalmers et al. (2016) as implemented in the *mirt* (Chalmers, 2012) package version 1.30. The results showed that there was approximately 0.68% (0.40%, 1.17%) average absolute difference in test response curves (integrated over a theta range of -4.00 to 4.00) and approximately -0.06 (0.04, -0.16, $p = 0.21$) raw score bias on average. Both of these scores indicate minor overall impact of DIF at the aggregated score level. Plotting the signed differential test functioning suggested that there was a minor non-negligible DIF in the -2.00 to -0.50 theta range and 1.80 to 3.60 theta range, with a maximum of 0.29 and -0.65 raw score bias in these two ranges, respectively. These results can be obtained from the second author.

Practitioners should exercise proper care and consider appropriate decision criteria when using these two instruments in South African workplaces and should not limit decisions solely to scores on these two instruments. More specifically, a comprehensive mental health assessment process is required, of which these measuring instruments can form a part. As a cautionary note, psychologists in the workplace should not utilise these instruments to make individual conclusive diagnoses of depression and anxiety, as only practitioners who are within their scope of practice (e.g. psychiatrists) may do so. However, the instruments can potentially be used to measure prevalence within a group or population as a whole, and also to potentially identify individuals who are at risk for they should be referred to an appropriate mental health practitioner for formal diagnosis and treatment.

Limitations and recommendations

Overall the study results support the validity of the CESD-R and GAD-7 items from the perspective of DIF for African and white working adults in this sample group. However, our results should be interpreted with caution. The sample size was quite small and most of the participants scored relatively low on these two instruments. As previously alluded to in this article, this can affect the precision of the parameter estimates, especially at the upper end of the latent distributions, making it difficult to determine the overall magnitude of the DIF (Chalmers et al., 2016). It is also possible that our relatively small sample size lacked the necessary power to detect statistically significant DIF although we attempted to correct for this by using a less strict *p*-value for determining DIF (Scott et al., 2009). Researchers should not interpret the results obtained in this sample as a definitive conclusion on DIF in the CESD-R and GAD-7 for African and white working adults. There is some evidence that DIF results generalise across multiple samples from the same population (Hamzeh, 2004) although it is not clear if our results will hold for these instruments when using different sample groups. Our results, however, serve as a useful starting point for the detection of DIF in the CESD-R and GAD-7 and we are optimistic that future studies can build on the promising results of our study. We also suggest that qualitative studies are needed (Sireci, 2011) to help determine the source of the bias in the items we identified as potentially problematic and to suggest improvements of the CESD-R and GAD-7 items for the South African context.

Conclusion

This study investigated DIF of the CESD-R and GAD-7 instruments in a South African sample of working persons and achieved its objective. Overall the results showed minor evidence for DIF in these instruments, although there was some non-negligible DIF in some items and at certain ranges of the theta distribution. This was especially true for the upper end of the theta distribution for the GAD-7. We believe that this study contributes to further research on the reliability and validity of these two instruments in the South African context, thereby enabling researchers and organisations to investigate mental health in detail with the required scientific rigour.

Acknowledgements

The first author (C.H.) extends their appreciation to the master's students for collecting the data used in this study.

Competing interests

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

Authors' contributions

Data were obtained from students supervised by C.H., who was also responsible for conceptualising and writing the manuscript. B.M. conducted conducted and interpreted the statistical analysis and assisted in writing the manuscript.

Funding information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability statement

Data used in this article can be obtained from the first author (C.H.).

Disclaimer

The views expressed in this article are the authors' own and do not reflect the official position of their institution.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edn., Text Revision). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2005). *Diagnostic and statistical manual of mental disorders* (4th edn.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th edn.). Washington, DC: American Psychiatric Association.

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

Beard, C., & Björgvinsson, T. (2014). Beyond generalized anxiety disorder: psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. *Journal of Anxiety Disorders, 28*(6), 547–552. https://doi.org/10.1016/j.janxdis.2014.06.002

Bender, A., & Farvolden, P. (2008). Depression and the workplace: A progress report. *Current Psychiatry Reports, 10*(1), 73–79. https://doi.org/10.1007/s11920-008-0013-6

Berger, M., & Tutz, G. (2015). Detection of uniform and nonuniform differential item functioning by item-focused trees. *Journal of Educational and Behavioral Statistics, 41*(6), 559–592. https://doi.org/10.3102/1076998616659371

Bezuidenhout, D., & Henn, C. M (2019). *Investigating the psychometric properties of the Generalized Anxiety Disorder Scale-7 (GAD-7) in a non-psychiatric sample of working people.* Manuscript in preparation.

Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 46*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software, 39*(8), 1–30.

Claassens, H. (2018). *Job resources as moderators of the relationship between job demands and well-being.* Unpublished master's dissertation. University of Johannesburg, Johannesburg, South Africa.

Crane, P. K., Gibbons, L. E., Jolley, L., & Van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Medical Care, 44*(11), S115–S123. https://doi.org/10.1097/01.mlr.0000245183.28384.ed

Carleton, R. N, Thibodeua, M. A., Teale, M. J. N., Welch, P. G., Abrams, M. P., Robinson, T., & Asmundson, G. J. G. (2013). The Center for Epidemiologic Studies Depression Scale: A review with a theoretical and empirical examination of item content and factor structure. *PLoS One, 8*(3), e58067. https://doi.org/10.1371/journal.pone.0058067

Chalmers, R. P., Counsell, A., & Flora, D. B. (2016). It might not make a big DIF: Improved differential test functioning statistics that account for sampling variability. *Educational and Psychological Measurement, 76*, 114–140. https://doi.org/10.1177/0013164415584576

De Beer, M. (2004). Use of differential item functioning (DIF) analysis for bias analysis in test construction. *South African Journal of Industrial Psychology, 30*(4), 52–58. https://doi.org/10.4102/sajip.v30i4.175

De Sá Junior, A., Liebel, G., De Andrade, A. G., Andrade, L. H., Gorenstein, C., & Wang, Y. -P. (2019). Can gender and age impact on response pattern of depressive symptoms among college students? A differential item functioning analysis. *Frontiers in Psychiatry, 10*(50). https://doi.org/10.3389/fpsyt.2019.00050

Donker, T., Van Straten, A., Marks, I., & Cuijpers, P. (2011). Quick and easy self-rating of Generalized Anxiety Disorder: Validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. *Psychiatry Research, 188*, 58–64. https://doi.org/10.1016/j.psychres.2011.01.016

Eaton, W. W., Smith, C., Ybarra, M., Muntaner, C., & Tien, A. (2004). Center for Epidemiologic Studies Depression Scale: Review and revision (CESD and CESD-R). In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment. Volume 3: Instruments for Adults* (3rd edn., pp. 363–377). Mahwah, NJ: Lawrence Erlbaum.

Evans-Lacko, S., Koeser, L., Knapp, M., Longhitano, C., Zohar, J., & Kuhn, K. (2016). Evaluating the economic impact of screening and treatment for depression in the workplace. *European Neuropsychopharmacology, 26*(6), 1004–1013. https://doi.org/10.1016/j.euroneuro.2016.03.005

Gamerman, D., Gonçalves, F. B., & Soares, T. M. (2018). Differential item functioning. In W. J. Van Der Linden (Ed.), *Item response theory: Applications* (Vol. 3, pp. 67–86). Boca Raton, FL: Taylor & Francis.

García-Campayo, J., Zamorano, E., Ruiz, M. A., Pardo, A., Pérez-Páramo, M., López-Gómez, V., et al. (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

Gravetter, F. J., & Forzano, L. A. B. (2015). *Research methods for the behavioral sciences.* Belmont, CA: Cengage Learning.

Hagquist, C., & Andrich, D. (2017). Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health and Quality of Life Outcomes, 15*, 1–8. https://doi.org/10.1186/s12955-017-0755-0

Hamzeh, D. (2004). Stability of differential item functioning over a single population in survey data. *The Journal of Experimental Education, 72*(3), 181–193. https://doi.org/10.3200/JEXE.72.3.181-193

Harnois, G., & Gabriel, P. (2002). *Mental health at work: Impact, issues and good practices.* Geneva: World Health Organization and International Labor Organisation, Department of Mental Health and Substance Dependence (MSD).

Haslam, C., Atkinson, S., Brown, S. S., & Haslam, R. A. (2005). Anxiety and depression in the workplace: Effects on the individual and organisation (a focus group investigation). *Journal of Affective Disorders, 88*(2), 209–215. https://doi.org/10.1016/j.jad.2005.07.009

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*(2), 339–344. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3191537/pdf/nihms327590.pdf.

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

Hope, D., Adamson, K., McManus, C., Chis, L., & Elder, A. (2018). Using differential item functioning to evaluate potential bias in a high stakes postgraduate knowledge based assessment. *BMC Medical Education, 18*(64), 1–7. https://doi.org/10.1186/s12909-018-1143-0

International Test Commission. (2013). *ITC guidelines on test use.* Retrieved from https://www.intestcom.org/files/guideline_test_use.pdf

Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type-I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education, 14*, 329–349. https://doi.org/10.1207/S15324818AME1404_2

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.

Joyce, S., Modini, M., Christensen, H., & Mykletun, A. (2016). Workplace interventions for common mental disorders: A systematic meta-review. *Psychological Medicine, 46*(4), 683–697. https://doi.org/10.1017/S0033291715002408

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (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

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

Michas, M. (2018). *Validation of the CESD-R in a non-clinical sample in the South African workplace*. Unpublished master's dissertation. University of Johannesburg, Johannesburg, South Africa.

Michas, M., & Henn, C. (2019). *Validation of the CESD-R in a non-clinical sample of working persons*. Manuscript preparation.

Mykletun, A., & Harvey, S. B. (2012). Prevention of mental disorders: A new era for workplace mental health. *Occupational and Environmental Medicine, 69*(12), 868–869. https://doi.org/10.1136/oemed-2012-100846

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

Parkerson, H. A., Thibodeau, M. A., Brandt, C. P., Zvolensky, M. J., & Asmundson, G. J. (2015). Cultural-based biases of the GAD-7. *Journal of Anxiety Disorders, 31*, 38–42. https://doi.org/10.1016/j.janxdis.2015.01.005

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401. https://doi.org/10.1177/014662167700100306

R Core Team (2018). R: *A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Republic of South Africa. (1974). *Health Professions Act Number 56 of 1974*. Pretoria: Government Press.

Republic of South Africa. (1998). *Employment Equity Act Number 55 of 1998*. Pretoria: Government Press.

Revelle, W. (1979). Hierarchical cluster-analysis and the internal structure of tests. *Multivariate Behavioral Research, 14*, 57–74. https://doi.org/10.1207/s15327906mbr1401_4

Revelle, W., & Zinbarg, R. E. (2008). Coefficients alpha, beta, omega and the glb: Comments on Sijtsma. *Psychometrika, 74*, 145. https://doi.org/10.1007/s11336-008-9102-z

Rutter, L. A., & Brown, T. A. (2017). Psychometric properties of the Generalized Anxiety Disorder Scale-7 (GAD-7) in outpatients with anxiety and mood disorders. *Journal of Psychopathology and Behavioral Assessment, 39*(1), 140–146. https://doi.org/10.1007/s10862-016-9571-9

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monographs, 34*(4, Pt. 2).

Scott, N. W., Fayers, P. M., Aaronson, N. K., Bottomley, A., De Graeff, A., Groenvold, M., … Sprangers, M. A. G. (2009). A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. *Journal of Clinical Epidemiology, 62*, 288–295. https://doi.org/10.1016/j.jclinepi.2008.06.003

Sekatane, D. (2018). *Personality as moderator in the relationship between anxiety and occupational health outcomes*. Unpublished master's dissertation. University of Johannesburg, Johannesburg, South Africa.

Sireci, S. G. (2011). Evaluating test and survey bias across languages and cultures. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 216–243). New York, NY: Cambridge University Press.

Sousa, T. V., Viveiros, V., Chai, M. V., Vicente, F. L., Jesus, G., Carnot, M. J., … Ferreira, P. L. (2015). Reliability and validity of the Portuguese version of the Generalized Anxiety Disorder (GAD-7) scale. *Health and Quality of Life Outcomes, 13*(1), 1–8. https://doi.org/10.1186/s12955-015-0244-2

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

Steyn, R., & De Bruin, G.P. (2018). Investigating the validity of the Human Resource Practices Scale in South Africa: Measurement invariance across gender. *SA Journal of Human Resource Management, 16*(1), 1-10. https://10.4102/sajhrm.v16i0.1038

Tan, L., Wang, M. -J., Modini, M., Joyce, S., Mykletun, A., Christensen, H., & Harvey, S. B. (2014). Preventing the development of depression at work: A systematic review and meta-analysis of universal interventions in the workplace. *BMC Medicine, 12*(74), 1–11. https://doi.org/10.1186/1741-7015-12-74

Tiirikainen, K., Haravuori, H., Ranta, K., Kaltiala-Heino, R., & Marttunen, M. (2019). Psychometric properties of the 7-item Generalized Anxiety Disorder Scale (GAD-7) in a large representative sample of Finnish adolescents. *Psychiatry Research, 272*, 30–35. https://doi.org/10.1016/j.psychres.2018.12.004

Tsebe, D. (2018). *Depression in the workplace, burnout and work engagement: Personality as moderator*. Unpublished master's dissertation. University of Johannesburg, Johannesburg, South Africa.

Van Dam, N. T., & Earleywine, M. (2011). Validation of the Center for Epidemiologic Studies Depression Scale-Revised (CESD-R): Pragmatic depression assessment in the general population. *Psychiatry Research, 186*(1), 128–132. https://doi.org/10.1016/j.psychres.2010.08.018

Van De Vijver, F. J. R., & Leung, K. (2011). Equivalence and bias: A review of concepts, models, and data analytic procedures. In D. Matsumoto & F. J. R. Van De Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 17–45). New York, NY: Cambridge University Press.

Walsh, T. (2014). *The conceptualization of depression and acculturative stress among Latino immigrants: A comparison of scores from Non-Hispanic Whites and persons of Mexican origin on the Center for Epidemiologic Studies Depression Scale-Revised (CESD-R).* Doctoral dissertation. The University of North Carolina at Chapel Hill, Chapel Hill, NC.

World Health Assembly. (2012). *Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level: Report by the Secretariat*. Retrieved from https://apps.who.int/iris/handle/10665/78898.

World Health Organization. (2014). *Mental health: A state of well-being.* Retrieved from https://www.who.int/features/factfiles/mental_health/en/.

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

Zumbo, B. D. (1999). *A Handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.