AOSIS

The structural validity of the Experience of Work and Life Circumstances Questionnaire (WLQ)

CrossMark
click for updates

Authors:
Pieter Schaap[1]
Esli Kekana[1]

Affiliations:
[1]Department of Human Resources Management, University of Pretoria, South Africa

Corresponding author:
Pieter Schaap,
pieter.schaap@up.ac.za

Orientation: Best practice frameworks suggest that an assessment practitioner's choice of an assessment tool should be based on scientific evidence that underpins the appropriate and just use of the instrument. This is a context-specific validity study involving a classified psychological instrument against the background of South African regulatory frameworks and contemporary validity theory principles.

Research purpose: The aim of the study was to explore the structural validity of the Experience of Work and Life Circumstances Questionnaire (WLQ) administered to employees in the automotive assembly plant of a South African automotive manufacturing company.

Motivation for the study: Although the WLQ has been used by registered health practitioners and numerous researchers, evidence to support the structural validity is lacking. This study, therefore, addressed the need for context-specific empirical support for the validity of score inferences in respect of employees in a South African automotive manufacturing plant.

Research design, approach and method: The research was conducted using a convenience sample (N = 217) taken from the automotive manufacturing company where the instrument was used. Reliability and factor analyses were carried out to explore the structural validity of the WLQ.

Main findings: The reliability of the WLQ appeared to be acceptable, and the assumptions made about unidimensionality were mostly confirmed. One of the proposed higher-order structural models of the said questionnaire administered to the sample group was confirmed, whereas the other one was partially confirmed.

Practical/managerial implications: The conclusion reached was that preliminary empirical grounds existed for considering the continued use of the WLQ (with some suggested refinements) by the relevant company, provided the process of accumulating a body of validity evidence continued.

Contribution/value-add: This study identified some of the difficulties that assessment practitioners might face in their quest to comply with South Africa's regulatory framework and the demands of contemporary test validity theory.

Introduction

Psychometric assessment serves to reach an understanding of and to describe underlying factors that drive individuals to display certain behaviours (Foxcroft & Roodt, 2013). The results from such assessments can enable psychologists to understand a person holistically and can potentially enable professionals to predict how a person might behave or react in different contexts (Carter, 2011). These results provide rich information about why people might behave in a certain way (Moerdyk, 2009). Using psychometric assessment instruments as part of a selection or development process may provide rich information that can aid organisations in identifying suitable candidates based on their aptitude, values, motivational drivers and personality (Carter, 2011). Although these instruments can be useful, their use holds some potential risks to companies in South Africa, given the country's socio-political history (Foxcroft & Roodt, 2013).

The Republic of South Africa (RSA) has a long history of discrimination against non-white groups. As part of the move towards nationwide transformation, the *Employment Equity Act 55 of 1998* (EEA) (RSA, 1998) was introduced. Its main objectives are promoting equal opportunity in the workplace and prohibiting unfair discrimination (RSA, 1998). Section 8 of the EEA relates to psychological testing in the workplace. It states that 'Psychological Testing and other similar assessments are prohibited unless the test or assessment being used – (a) has been scientifically shown to be valid and reliable, (b) can be fairly applied to all employees, and (c) is not biased

against any employee or group' (RSA, 1998). One of the most recent amendments of the EEA relates to Section 8(d), which states that psychological or similar assessments need to be 'certified by the Health Professions Council of South Africa (HPCSA) established by Section 2 of the Health Professions Act, 1974 (Act No. 56 of 1974), or any other body which may be authorised by law to certify those tests or assessments' (RSA, 2013).

In terms of the EEA and the *Health Professions Act of 1974*, the test user is responsible for ensuring that measures selected for use are reliable, valid, culturally fair and unbiased (Van de Vijver & Rothman, 2004). The intention of these laws is good, but their application appears to be a challenge to practitioners (Kubiszyn *et al.*, 2000; Theron, 2007). The current study explored the structural validity of an assessment tool, the Experience of Work and Life Circumstances Questionnaire (WLQ) (classified by the HPCSA as a psychological test), when used to assess employees of an automotive assembly plant in South Africa. The study was done against the background of the existing regulatory frameworks in South Africa. The practical challenges the South African regulatory frameworks pose to the use of tests, in particular when looking at the legal compliance of these tests through the lens of contemporary test validity theory, were pointed out. The value of a context-specific structural validity study of the WLQ as a preliminary form of validity evidence in increasing the likelihood that the use of the measure would stand up to legal scrutiny, was demonstrated.

The onus has always been on psychologists to ensure that any instrument they use is aligned with the principles of best practice and ethical conduct and that it meets legal requirements (Department of Health, 2006). However, Section 8 of the EEA may be considered inherently problematic in respect of the clause that tests are prohibited unless scientifically shown to be reliable and valid. There are strong arguments against the notion that a psychological measure can be considered compliant with the EEA at any point (Theron, 2007). It is well recognised that judgements about validity are more appropriately defined as 'unitary' and 'evolving' on a continuum of evidentiary support for intended score inferences and use and not as dichotomous (the claim that tests are either valid or non-valid) (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA & NCME], 2014; Cizek, 2015; Kane, 2015). The EEA is not clear on specifically what scientific evidence on the continuum of evidentiary support would be deemed sufficient to regard the WLQ as reliable and valid to be used for employees, for instance, of an automotive assembly company. Test scores from a specific test may or may not have equal validity if a test is used for decision making in different organisations whose employees, cultures, processes, goals and performance measures differ (Guion & Highhouse, 2011). This lack of clarity leaves test users with a measure and a criterion for compliance that are undefined and, therefore, open to interpretation, for instance, in a court

of law. Furthermore, there are no recorded court judgements to guide users on particular measures and criteria for compliance. The inclusion of clause (d) in the EEA may also lead to the erroneous interpretation that all tests certified or classified by the HPCSA as being psychological in nature are legally compliant and therefore meet validity, reliability and fairness requirements according to the stipulations of the EEA. Such conclusions would be inherently flawed from a substantive and process point of view and are not in line with contemporary theories on valid test score inferences and use (AERA *et al.*, 2014; Cizek, 2015; Guion & Highhouse, 2011; Kane, 2015).

The mandate of the HPCSA is to classify tests as being psychological or non-psychological after having considered the relevant and prescribed information, including reliability and validity evidence, supplied by the test's publishers and the evaluations of independent reviewers (HPCSA, 2010). However, this process cannot claim to provide conclusive validity evidence on a measure in terms of the principles of contemporary theories of test validation (Cizek, 2015; Guion & Highhouse, 2011; Kane, 2015). The HPCSA's list of classified assessment instruments (Form 207) was last updated in 2010 and still includes instruments that were classified on the basis of research evidence from the Apartheid era (HPCSA, 2014a). The HPCSA Notice 93 of 2014 indicates which instruments have been classified and reviewed, which ones have been classified but not reviewed and which ones are still under development (HPCSA, 2014a). The WLQ is one of the instruments that appear as classified on Form 207 but has not undergone the review process.

The objectivity and authenticity of the test reliability and validity information published by researchers and test publishers may sometimes be questionable (McDaniel, Rothstein & Whetzel, 2006). Guion and Highhouse (2011) point out that a critical stance should be taken towards validity evidence – both confirmatory and non-confirmatory findings should be reported in a balanced way and plausible alternative hypotheses should be investigated. Confirmation bias in studies published in psychology is a growing concern, and pressure mounts for increased academic research output over shorter periods (Wagenmakers, Wetzels, Borsboom, Van der Maas, & Kievit, 2012). These authors and Ferguson and Heene (2012) argue that it is entirely understandable that researchers prefer to select analysis strategies that confirm their pre-formulated hypotheses and avoid ones that do not, as null results are difficult to interpret and therefore likely to elicit negative comment from reviewers. It is hard to imagine that test distributors, who are in the business of selling tests, would report non-confirming, product-specific research results in the test manuals presented to the reviewers appointed by the Psychometrics Committee of the HPCSA.

It should be clear from the arguments presented that the WLQ's continued use in the company under study may ultimately depend on validity evidence in context of use for the following reasons:

- Relying on published research results as sources of validity evidence, whether they have been taken up in a test manual or in other publications, may be problematic. Judgements about the validity of a specific instrument for a specific context of use which is based on a body of published evidence of a generic nature may be seen as challengeable in most instances, as the evidence may be biased or not relevant to the context in question.
- Tests that are classified or certified cannot be considered an endorsement of the test's validity in terms of the requirements of the EEA, and more so if the test has been classified but not reviewed as is the case with the WLQ.
- When test validity in terms of the EEA is challenged in court, it might be problematic to respond to it as there is no clarity on the body of validity evidence that would be deemed sufficient for a test to be considered reliable and valid. Therefore, the outcome of the case might hinge on context-specific validity evidence rather than on published evidence of a generic nature.

The HPCSA's policy on the classification of psychometric measures clearly states that test users should contribute to empirical studies on tests (HPCSA, 2010). One would assume that studies of this nature should be conducted in accordance with the guidelines of contemporary validity theories to ensure scholarly and professional recognition (AERA *et al.*, 2014). In addition, it may be argued that psychologists have an ethical responsibility to make their findings known in the interests of the general public and the profession.

Purpose and objectives

If the WLQ is to be used in the particular automotive manufacturing company, it is imperative that the relevant validity information is obtained for the purpose of interpreting and using the test scores of the employees of this company. Aside from the information in the WLQ manual (Van Zyl & Van der Walt, 1991) and in a later publication by Oosthuizen and Koortzen (2009), no other research that focuses specifically on the evaluation of the internal structural properties of the WLQ, could be found. In terms of contemporary validity theories, multiple forms of validity are required to establish the validity of test scores in a specific context (AERA *et al.*, 2014). However, the type of data available about the instrument at the time of this study allowed only for internal structural validity study. These can be considered a logical starting point in collecting a body of evidence to evaluate the suitability of the instrument for use in the relevant automotive manufacturing company.

The purpose of this study was thus to explore and evaluate the structural validity of WLQ scores obtained in respect of a sample of employees in the automotive assembly plant of an automotive manufacturing company in South Africa.

More specifically, the objectives of this study were to:

- explore and evaluate the assumptions of reliability and unidimensionality of the WLQ administered to a sample group of employees from a South African automotive manufacturing plant and
- determine the structural validity of the proposed higher-order structural measurement models of the WLQ administered to the sample group.

The main contribution of this study was to provide evidence pertaining to the structural validity of the WLQ scores in respect of the sample of employees and also to highlight the practical challenges of meeting the requirements of Section 8 of the EEA against the backdrop of contemporary theoretical perspectives on test validity.

Literature review
Contemporary validity theory

As pointed out earlier, the stipulations of Section 8 of the EEA do not appear to accommodate the fundamental principles of contemporary validity theory, an issue that poses a challenge. Practitioners are compelled to acknowledge contemporary validity theory principles if their research initiatives are to be recognised by scholars and professionals. The fundamental principles of contemporary validity theory (AERA *et al.*, 2014; Bornstein, 2011; Cizek, 2015; Kane, 2015; Messick, 1989; Newton & Shaw, 2015; Shepard, 1993) can be summarised as follows:

- There is general consensus that validity pertains to the intended inferences drawn from test scores or the interpretation of these test scores and not the tests themselves.
- There is consensus that judgements about validity are not dichotomous (for instance, that tests are valid or invalid) nor are they based on a set of correlation coefficients: validity is appropriately described along a continuum of evidentiary support for the intended score inferences and their use.
- Contemporary validity theorists largely reject the existence of diverse kinds of validities and favour the 'unitary' concept of validity according to which a collective body of evidence is required.
- There is agreement that validation is a process and not a one-time activity.
- The notion is accepted that inquiries into evidence related to test constructs, intended test use and the consequences of testing are different processes.

According to the most recent Standards for Educational and Psychological Testing (hereafter, 2014 Standards), 'validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests' (AERA *et al.*, 2014, p. 11). The types of validity evidence required according to these standards relate to test content representation and relevance, response processes (participants' responses to items and differences in groups' responses), internal structure (Cronbach's alpha and KR 20 reliability indices, factor structures, item statistics and differential item functioning), other variables (criterion validity, convergent and discriminant validity, experimental

and longitudinal studies) and the consequences of testing (the expected benefits of testing and possible unintended consequences) (Bornstein, 2011). The 2014 Standards suggest that convergence of the multiple forms of validity is required to establish the validity of test scores in a specific context. Therefore, validity is no longer considered inherent in distinct forms but is conceptualised as a 'unitary concept' in the sense that a collective body of validity evidence is required (Bornstein, 2011; Messick, 1989).

Factors that contribute to the need for ongoing validity research include changes in intended test populations, the evolution of theoretical constructs, the mode of administration and scoring, the context of use, and many other factors that could alter the conclusions about the validity of score inferences and their use (Cizek, 2015; Guion & Highhouse, 2011). Cizek (2015), Sireci (2015) and Kane (2015) distinguish clearly between evidence related to score inferences and evidence related to the context of a test's use. It is argued that these endeavours to provide evidence are separate and should be treated separately. Nevertheless, in some circumstances, the anticipated consequences or benefits of test use can confirm or disconfirm score inferences (Cizek, 2015).

Comprehensive evidence supporting score inferences is a prerequisite, but it is not a factor dictating how and where tests are used. However, from a professional perspective, it is unthinkable to use tests in any given context without sufficient evidence supporting score inferences. Hence, the 2014 Standards (p. 16) have subsumed 'evidence-based internal structure' and 'evidence based on relations with other (external) variables' under the overarching heading of 'Evidence based on hypothetical relationships between variables (internal and external)'. The underlying message is that score inference validity research should be theory driven. This implies that the interpretation of relationships between variables should be guided by grounded theoretical propositions (Cizek, 2015; Messick, 1989). The internal structure of a measure consists of the response patterns of the test takers (the data) relating to items included in the measure. The data should support the a priori theoretical propositions relevant to the internal structure of a measure, better known as its structural or factorial validity.

The next section describes the WLQ from a theoretical perspective and considers related propositions and hypotheses for the purposes of score inference validation.

The Experience of Work and Life Circumstances Questionnaire theoretical background

The WLQ is based on the model of stress of Cox and MacKay (1981), which, in turn, is rooted in Lazarus's definition of stress, according to which a person's ability to overcome stressful situations depends on the perception that the person has of the stressor and of his or her ability to overcome the stressor (Lazarus & Folkman, 1984; Van Zyl & Van der Walt, 1991). According to Robbins and Judge (2015),

stress is 'an unpleasant psychological process that occurs in response to environmental pressures. Stress is a dynamic condition in which an individual is confronted with an opportunity, demand, or resources related to what an individual desires and for which the outcome is perceived to be both uncertain and important'. However, a generally accepted definition of the term stress appears to be elusive as it can be given from different perspectives, for instance, from the point of view of the engineering (environmental), physiological or psychological sciences. The influential work of Cox and Lazarus on work-related stress has a psychological emphasis in that they conceive the experience of stress as being more than a mere stimuli and response process; it is a dynamic process that occurs as an individual interacts with their environment (Cox, 1978; Lazarus & Folkman, 1984, as cited in Mark & Smith, 2008). Moreover, stress is seen as a condition that arises when real or perceived demands exceed real or perceived abilities to deal with the demand, resulting in a disturbance of employees' psychological and physiological equilibrium (Colligan & Higgins, 2005). Perception plays an important role in the experience of stress and determines whether a situation is perceived as being threatening. Perception of a situation occurs at a cognitive level of appraisal, and a person's emotional, cognitive, behavioural and physiological responses to stress depend on the characteristics of the stressor, the resources available to counter the stressor and the individual's personal characteristics (Colligan & Higgins, 2005; Cox, 1978).

Stress responses may be positive (normal stress or eustress) and lead to healthy consequences (e.g. high energy levels, mental alertness, optimistic outlook and self-motivation) up to an optimum point. However, after prolonged exposure to stress, a negative connotation (stressfulness, distress) arises leading to unhealthy consequences (e.g. psychological disorders, such as depression and burnout, and psychosomatic disorders, such as heart disease, strokes and headaches) (Cox, 1978; Nelson & Quick, 2000). Negative consequences of work stress may manifest as physiological symptoms (headaches, high blood pressure, heart disease), psychological symptoms (anxiety, depression, decreased job satisfaction) and behavioural symptoms (low productivity, absenteeism and high turnover) (Robbins & Judge, 2015).

The focus of the WLQ is on the distress component of stress that arises from an accumulation of stressors or demands that are perceived by an employee to exceed the available resources and abilities to cope with these demands (Hu, Schaufeli & Taris, 2011). Van Zyl and Van der Walt (1991) adopted the model of stress developed by Cox and MacKay (1981) as a theoretical foundation for understanding the WLQ measurement (Cox, 1978, p. 19). The interpretation of the model by Van Zyl and Van der Walt (1991) consists of four phases and is summarised as follows:

- Phase 1: The individual's evaluation of internal and external demands to determine their status (harm, threat and challenge).

- Phase 2: The individual's evaluation of his or her ability to cope with the demand perceived in Phase 1.
- Phase 3: The individual's emotional, cognitive and behavioural response to the identified stressor.
- Phase 4: The effect of this response on the individual.

According to Van Zyl and Van der Walt (1991), the WLQ suggests that the experience of stress (Phases 3 and 4 of the model) may be a product of various external and internal environmental sources (Phases 1 and 2 of the model). The literature supports the notion that stressors may have their origins outside or inside the work environment (Catellano & Plionis, 2006; Matteson & Ivansevich, 1987). The aim of the WLQ is to determine an employee's level of stress, as well as the sources of this stress from both outside and inside the work environment (Van Zyl & Van der Walt, 1991). When used for diagnostic purposes in a work environment, the WLQ can indicate whether high levels of stress can be attributed to environmental factors and whether these play a significant role.

The literature also supports the notion of a variety of potential sources of stress. Potential sources that have been identified by Robbins and Judge (2015) include environmental factors (economic, political and technological), organisational factors (task demands, role demands and interpersonal demands) and personal factors (family problems, work–family conflict, economic problems and personality). The environmental and personal factors identified by these authors are conceptualised in the WLQ model as external to the work environment, whereas the organisational factors are conceptualised as internal to the work environment. Stressors internal to the work environment have been well researched in the literature. Mark and Smith (2008) refer to five categories of workplace stress, which include factors unique to the job (work hours, work autonomy, meaningfulness of work, physical work environment and workload), employees' role in the organisation (responsibility, oversight and role ambiguity), career development (promotion, career opportunities, mergers and acquisitions and technological changes), interpersonal relationships in the workplace (problematic relationships, harassment and workplace bullying) and overall organisational structure and climate (organisational communication patterns, management style and participation in decision making). The role of physical environmental pressures in the workplace on work stress have been recognised by Matteson and Ivansevich (1987) and Arnold, Cooper and Robertson (1995) as potential sources of stress. These pressures relate directly to unpleasant work conditions, physical design, noise levels and hygiene. Hristov *et al*. (2003) contend that a discrepancy between work performance and remuneration policies is a potential source of work stress, especially when combined with a high workload. Remuneration and employee benefits have an impact on the social status, level of expectations and quality of life of an employee and their families, especially under economically challenging conditions (Fontana, 1994; Lemanski, 2003).

The Experience of Work and Life Circumstances Questionnaire measurement model

The first part of the WLQ consists of 40 items (Scale A) that measure an individual's experiences related to stress, which include despondency, self-confidence, conflict with others, anxiety, worries, isolation, guilt, control and uncertainty (Van Zyl & Van der Walt, 1991). The second part of the questionnaire consists of 76 items that aim to determine whether stress arises as a result of the individual's personal environment (Scale B) or work environment with regard to circumstances or missed expectations (Scale C) (Van Zyl & Van der Walt, 1991).

The WLQ manual lists the contributing factors that cause stress outside the work environment but does not provide a clear definition of the construct that underlies Scale B. Some of these factors are family problems, financial circumstances, phase of life, general economic situation in the country, changing technology, facilities at home, social situations, status, health, background, effect of work on home life, transport facilities, religious life, political views, availability of accommodation and recreational activities (Van Zyl & Van der Walt, 1991).

The work-related causes of stress taken up in the WLQ originated from the Job Description Index (Smith & Kendall, 1963) and research done by Ludik (1988). In the work environment, the unfulfilled expectations (Scale C) that cause stress consist of the following subscales: organisational functioning (Scale C1), task characteristics (Scale C2), physical work environment (Scale C3), career matters (Scale C4), social matters (Scale C5) and remuneration, fringe benefits and to function under a just personnel policy on remuneration (Scale C6) (Van Zyl & Van der Walt, 1991). Although the WLQ manual lists and describes the factors that cause stress in the work environment, it does not provide a clear definition of the constructs underlying Scale B and Scale C, or any of the subscales. The subscales can be described as follows:

- Organisational functioning pertains to an employee's contribution to decision making and strategy creation, trust in the supervisor or manager, effective organisational structure, positive management climate, recognition and degree to which open communication can take place between an employee and a supervisor or manager.
- Task characteristics have to do with the extent to which employees can control their work, the degree to which employees are rewarded for work challenges they meet, the quality of instructions received for a task, the level of autonomy, reasonable deadlines, whether the work is enough to keep employees busy and whether the variety of work is sufficient.
- The physical work environment includes (in)adequate lighting, the temperature of the work area, the availability of proper office equipment, the cleanliness of the work area and the distance to and condition of the bathrooms.
- Career opportunities pertain to an individual's expectations regarding further training, use of skills and talents, career development and job security.

- Social matters in the workplace relate to an employee's experience of social interaction in the work environment and include expectations with regard to high status in an employee's job, positive relations with the manager/supervisor and colleagues, and reasonable social demands.

- Remuneration expectations refer to a fixed, commission-based or piece-rate salary and to function under a just personnel policy on remuneration. Fringe benefits include any tangible or intangible rewards that a company offers its employees. The financial aspect of work can be a source of stress because it is connected with a person's livelihood and/or identity.

Hattie (1985) argues that in order to make correct inferences from a scale score, the assumptions of unidimensionality, internal consistency or homogeneity of the item meaning should hold. It has rightly been noted that the probability that scales would be purely unidimensional in behavioural studies is quite small as the constructs under investigation are complex and may need to include heterogeneous content to represent the construct adequately (Reise, Moore & Haviland, 2010). Based on the arguments presented, the following hypothesis was formulated for this study:

Hypothesis 1: Each scale of the WLQ is essentially unidimensional.

Reise *et al.* (2010) argue that essential unidimensionality reflects contexts where scores represent a single common construct even when data are multidimensional, provided that the observed multidimensionality can reasonably be attributed to negligible item content parcels (nuisance factors). Oosthuizen and Koortzen (2009) report factor structures for the WLQ based on Principal Component Analyses using Kaiser's eigenvalues larger than unity as the factor extraction criterion. They extracted multiple factors for each scale; however, other than describing the factor content, they do not indicate the significance and plausibility of the identified factors or comment on the extent to which the WLQ scales' dimensionality is supported by the data. It is commonly accepted that the Kaiser's eigenvalue criterion leads to an over-extraction of factors consisting mostly of nuisance or non-significant factors. The WLQ manual contains no information on factor analytic models or structures that can provide a basis for formulating an a priori hypothesis in respect of scale dimensionality for the purposes of this study, so this had to be investigated using exploratory methods.

Van Zyl and Van der Walt (1991, p. 5) suggest that people's experience of stress may be related to circumstances in their work environment (internal causes of stress) or circumstances outside the work environment (external causes of stress). The high inter-correlations between the internal causes of stress subscales in the WLQ suggest a higher-order factor, namely that of internal causes of stress. The scale inter-correlation matrix as described by Van Zyl and Van der Walt (1991, p. 24) suggests significant inter-correlations between the scales of internal causes of stress and external causes of stress. The WLQ manual provides a priori substantiated theoretical propositions relevant to the internal structure of the respective models (Cox, 1978; Cox & McKay, 1978). Moreover, the model

suggested by Van Zyl and Van der Walt (1991) is well entrenched in work stress theory (Cartwright & Cooper, 1997; Nelson & Quick, 2000; Mark & Smith, 2008; Robbins & Judge, 2015). The structural model presented in Figure 1 can therefore be constructed based on the argument relating to the scale inter-correlation matrix put forward by Van Zyl and Van der Walt (1991) and the literature substantiation provided in the WLQ manual. Because of the confusion, misunderstanding and disagreement in the literature regarding the use of the term 'cause' or 'causal modelling' (Schreiber, Nora, Stage, Barlow, & King, 2006, p. 4), the researchers of this study adopted the recommendation of Schreiber *et al.* (2006) to simply refer to direct or total effects among the observable and latent variables as suggested by theoretical and empirical propositions.

The following hypothesis relating to a hypothesised WLQ model was, therefore, tested for the purposes of this study:
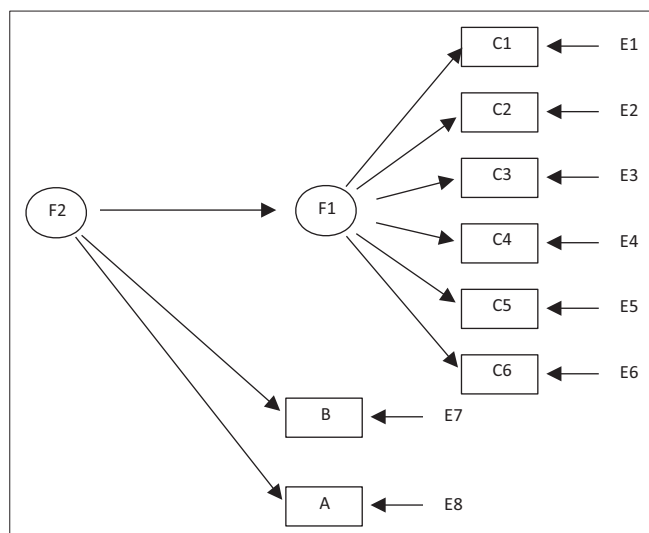
Hypothesis 2: The WLQ higher-order structural model is supported by the sample data, which suggests that variable A (Scale A) is directly affected by F1 (Scale C) and that A is directly affected by B (Scale B).

According to Van Zyl and Van der Walt (1991, pp. 2, 20), the WLQ as a whole represents the construct of work stress (higher-order latent variable) that comprises a joint measure of corresponding fields, which, in turn, consists of the level of stress experienced (Scale A), causes that arise outside the work situation (Scale B) and causes that originate within the work situation (subscales C1 to C6). Theoretical substantiation supporting this proposition put forward by Van Zyl and Van der Walt (1991) has been well articulated (Cartwright & Cooper, 1997; Cox, 1978; Cox & McKay, 1978; Mark & Smith, 2008; Nelson & Quick, 2000; Robbins & Judge, 2015). The proposition is further supported on the strength of the observed inter-correlations between the WLQ subscales reported in the WLQ manual. Based on the description provided by Van Zyl and Van der Walt (1991, p. 20), the structural model presented in Figure 2 was constructed for the purposes of this research.
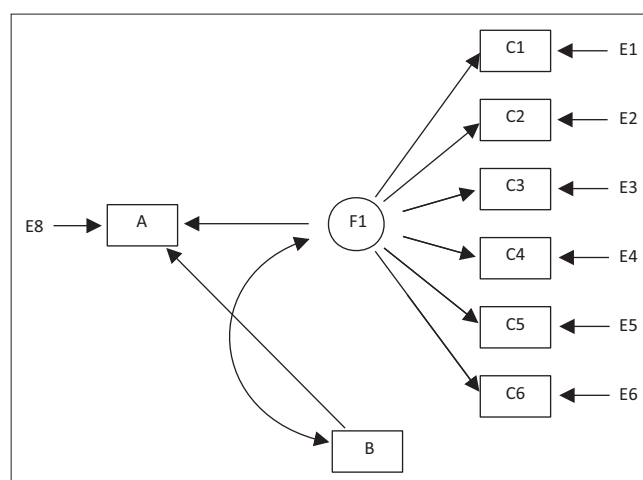
Source: Van Zyl, E.S., & Van der Walt, H.S. (1991). *Manual for the Experience of Work and Life Circumstances Questionnaire (WLQ)*. Pretoria: Human Sciences Research Council

FIGURE 1: Hypothesised higher-order structural model of the Experience of Work and Life Circumstances Questionnaire.

FIGURE 2: Hypothesised higher-order latent variable structural model of the Experience of Work and Life Circumstances Questionnaire.

The following hypothesis relating to a hypothesised WLQ model was tested for the purposes of this study:

> **Hypothesis 3:** The higher-order latent variable structural model of the WLQ is supported by the sample data and, therefore, valid inferences can be drawn from the variable scores.

In the next section of this paper, the research design followed to explore the plausibility of the three hypotheses is presented.

Research design

Research approach

In terms of the positivist paradigm followed in this research, a certain standard of realism based on observations is required for proven truth and knowledge (Babbie, 2013). The positivist paradigm forms the basis for most quantitative research because of its objective nature (Babbie, 2013).

Research method

As this study aimed to explore the psychometric properties of the WLQ (Straub, Boudreau & Gefen, 2004), the type of research embarked on was validation research. Furthermore, a non-experimental research design as recommended by Kumar (2011) was adopted. This study, which is rooted in a practical need to know, can be considered to be applied research. More specifically, a cross-sectional survey research design was followed in this study.

Research participants

A convenience sampling (non-probability sampling) method was used in this study to compile the 2014 data (Babbie, 2013). In accordance with this method, the researchers took advantage of the natural gathering of the participants to create the sample (Rempler & Van Ryzin, 2011). In the present study, the researcher approached the general managers (GMs) of 15 different departments of the company and asked their permission and

support to conduct the study. The GMs provided the researcher with adequate access to the sampling frame.

The sample consisted of 217 respondents. The average age of the respondents was 35.5, 72% of whom were below 40 years of age. With respect to gender, 70% ($n = 152$) were men and 30% ($n = 65$) were women. The majority indicated an African language as their first language (53.5%; $n = 116$), the second largest group indicated Afrikaans as their first language (30%; $n = 65$) and a minority indicated English (16.6%; $n = 36$) as their first language. Black Africans represented 54.4% ($n = 118$) of the sample, 38.7% ($n = 84$) was white, 2.8% ($n = 6$) was coloured and 4.1% ($n = 9$) was Indian. Most respondents had obtained a post-matric qualification: 35.5% ($n = 77$) had a post-matric certificate/diploma/degree, 47% ($n = 102$) had a postgraduate degree and 17.5% ($n = 38$) had completed education up to a matric/NQF4 level. Most of the sample group (53.5%; $n = 116$) had one or two dependants, 18.5% ($n = 40$) had three or more dependants and 28.1% ($n = 61$) had no dependants. With respect to length of service, 22.6% ($n = 49$) had less than 2 years, 26.7% ($n = 58$) had between 2 and 5 years and 50.7% ($n = 110$) had more than 6 years.

Practical rule of thumb put forth by Costello and Osbourne (2005) was used to determine the adequacy of the sample size for the purposes of this study. A subject-to-item ratio of between 5 and 10 respondents per item was the criterion that was used (Costello & Osbourne, 2005). The sample consisted of 217 respondents, thus providing a subject ratio of between 14 and 27 respondents per item, which was more than sufficient for the factor analyses done on each subscale.

The measuring instrument

The WLQ is a paper-and-pen–based questionnaire that consists of 3 sections, a total of 8 scales and 115 items (Van Zyl & Van der Walt, 1991). The items consist of relevant statements rated on a Likert scale with five options, where 1 is virtually never, 2 is sometimes, 3 is reasonably often, 4 is very often and 5 is virtually always (Van Zyl & Van der Walt, 1991).

The Kuder-Richardson 8 formula was used to determine the inter-item consistency with the standardisation of the WLQ in 1989 (Van Zyl & Van der Walt, 1991). Reliabilities of between 0.83 and 0.92 were obtained, which are regarded as satisfactory (Van Zyl & Van der Walt, 1991). Test–retest reliabilities (4 weeks apart) varied between 0.62 and 0.80 and were regarded as satisfactory (Van Zyl & Van der Walt, 1991).

Subject matter experts confirmed the face validity as an indicator of the content validity of the WLQ. According to Van Zyl and Van der Walt (1991, p. 22), the WLQ displays logical validity, which requires a careful definition in behavioural terms of the traits, analyses of all behavioural traits in parts where they are represented and the evaluation of item discrimination values.

Construct validity was confirmed using inter-test and intra-test scale inter-correlations that, according to Van Zyl and

Van der Walt (1991, p. 22), support the construct validity of the WLQ. The 16 Factor Questionnaire (Cattell, Eber & Tatsuoka, 1970), the Personal, Home, Social, and Formal Relations Questionnaire (Fouché & Grobbelaar, 1983), the Questionnaire on the Reaction to Demands in Life (which relates to cognitive and behavioural symptoms of stress), specifically developed by Van Zyl and Van der Walt (1991) to cross-validate the WLQ, were used for the purposes of the analyses. Other related studies provided further supporting evidence for the construct validity of the WLQ. Inter-test research conducted by Lamb (2009) indicated a significant linear relationship between job stress and two personality factors (emotional stability and open to experience). Research by Sardiwalla (2014) demonstrated strong positive relationships between levels of burnout and the WLQ's internal sources of stress, namely organisational functioning, task characteristics, career matters and external stressors. Van Zyl (2009) found that levels of work stress as measured by the WLQ were high amid uncertainty surrounding organisational change. Oosthuizen (2005) found that a sense of coherence and internal locus of control was related to a lower level of work stress, as measured by the WLQ. Roos and Van Eeden (2008) conducted a correlational analysis using the WLQ, the Motivation Questionnaire (Baron, Henley, McGibbon & McCarthy, 2002) and the Corporate Culture Lite Questionnaire (Davies, Phil & Warr, 2000), and found significant supporting evidence for the construct validity of the internal sources of stress.

Research procedure and ethical considerations

Ethical approval for conducting the study was given by the Post Graduate Committee of the Department of Human Resources Management at the University of Pretoria, and the approval of all stakeholders was obtained. The research proposal was approved by the automotive company on condition that the researcher protected the anonymity of the participants and the company. A formal consent form was signed by each participant after completing the WLQ. The consent included permission that the data may be used for research purposes and an agreement that the identity of each person will be protected.

The purpose of the study was explained to all stakeholders. E-mails requesting voluntary participation were sent to employees, and those who consented to participate signed informed consent forms. The researcher booked meeting rooms to be available at the times and on the dates participants would be available. A total of 217 employees from 15 different departments were assessed in a 3-month period (from 1 December 2013 to 28 February 2014). The participants took between 30 and 45 minutes to complete the WLQ. Each manager received feedback on the overall results for their departments, and trends in the results were pointed out.

Statistical analysis

The researchers gathered descriptive statistics on the WLQ scales, which included scale means, standard deviations and inter-scale correlations.

Cronbach's alpha score was used in this study to determine inter-item consistency. A Cronbach's alpha score above 0.90 is considered to be excellent, between 0.80 and 0.89 is good and between 0.70 and 0.79 is fair, but values below 0.70 are not acceptable (Dumont, Kroes, Korzilius, Didden & Rojahn, 2014).

Exploratory factor analysis (EFA) was chosen as it is commonly used in validation studies to identify the underlying factor structure of the instrument without it being influenced by the researcher (Hayton, Allen & Scarpello, 2004; Weng & Cheng, 2005). The highly recommended parallel analysis (PA) factor retention technique was also used, as it is considered the most unbiased approach to follow (Courtney, 2013). In addition, Cattell's scree test was used, as it is recommended that more than one technique be used to determine the number of factors to be retained. Kaiser's criterion was not used as research has indicated that this criterion tends to over-extract factors, which could make it more difficult to interpret the results (Courtney, 2013; Hayton *et al.*, 2004).

Principal Axis Factoring (PAF) in SPSS was used as the extraction method (Gaskin & Happell, 2014; Weng & Cheng, 2005), and the promax rotation method with a default kappa 4 value was used because the variables were interrelated. PAF analysis is preferred when the assumption of multivariate normality cannot be assumed, and there is no widely preferred method of oblique rotation as all tend to produce similar results (Costello & Osbourne, 2005).

The main objective of conducting an EFA was to determine the dimensionality of the WLQ because dimensionality affects the quality of the instrument as a whole (Smits, Timmerman & Meijer, 2012). The probability that the scales would be purely unidimensional was quite small, and based on the research findings of Oosthuizen and Koortzen (2009), it was expected that there would be some level of noise. If some of the scales are found to be multidimensional, a researcher can conduct bi-factor analysis (Wolff & Preising, 2005). In many cases where assessments indicate essential unidimensionality, evidence of multidimensionality is indicated at the same time (Reise *et al.*, 2010). The general or secondary factor is the overall factor that the items intend to measure, and the group factors are smaller factors within the general factor that are uncorrelated with one another but are correlated with the general factor (Wolff & Preising, 2005). If the general factor explains more than 60% of the total variance of the factor, it can be assumed that the factor is essentially unidimensional (Wolff & Preising, 2005). In the present study, the Schmid-Leiman analysis was used (Schmid & Leiman, 1957). Once the dimensionality of the scales had been determined, the second-order structural validity of the proposed model could be tested using a structural equation modelling (SEM) technique, namely Confirmatory Factor Analysis (CFA).

CFA was used together with the SEM software EQS 6.1 to test the structural validity of Model 1 and Model 2 of the WLQ, as

described earlier in Hypothesis 2 and Hypothesis 3 for the sample data. The maximum likelihood parameter estimation technique, coupled with the robust statistic that corrected for non-normality in the sample data, was used (Ullman, 2006).

Results

Experience of Work and Life Circumstances Questionnaire scale reliability and descriptive statistics

The WLQ scale alpha reliability coefficients and descriptive statistics are provided in Table 1. The reliability coefficients (presented in the matrix diagonal) obtained for the WLQ ranged from acceptable to very good. Scale A displayed excellent alpha reliability coefficients (0.95), whereas the coefficients of other scales showed good reliability and were as follows: B (0.82), C1 (0.82), C4 (0.84), C5 (0.82) and C6 (0.88). Scales C2 and C3 achieved a fair reliability score (0.70 and 0.79, respectively). Overall, the reliability score of the WLQ was good. All the reliability coefficients appeared to exceed the level of minimum acceptability and were fair to good.

Factor analyses

Results of exploratory factor analysis and bi-factor analyses

This section evaluates the results obtained from the EFA to determine whether the scales are unidimensional or multidimensional. The criteria used to determine dimensionality are based on Horn's PA, which determines how many factors should be retained (O'Connor, 2000) and the total variance explained by each factor (Lai, Crane & Cella, 2006). According to Lai *et al.* (2006), a minimum of 5% of the total variance should be explained by each factor that is retained. If the factor structure is good, the cumulative variance explained should be about 40% (Pallant, 2007).

As depicted in Figure 3, Scales C1, C3, C4 and C5 displayed evidence of unidimensionality (one factor accounted for at least 40% of the total variance). In the case of Scales A, B, C2 and C6, there was evidence of multidimensionality; therefore, a bi-factor analysis was conducted on these scales. A factor should consist of at least three or more items with a factor loading of 0.3 or higher, excluding items that cross-load. Figure 3 indicates a detailed view of each of the scales and its

factor loadings. Bi-factor analysis was conducted on the scales that appeared multidimensional, and the results are set out in Figure 4.

As indicated in Figure 4, Scale A's apparent multidimensionality could be ascribed to the presence of a general factor (62% of the variance was explained by this general factor). Thus, Scale A showed evidence of essential unidimensionality. Only two items (items A7 and A13) did not display salient loadings on the general factor.

Scale B appeared to contain problematic items. As demonstrated in Figure 3, Scale B's apparent multidimensionality could be explained by two factors. Figure 4 illustrates that the general factor explained 58% of the variance, whereas the group factors accounted for smaller portions of variance individually. This suggests that Scale B displayed some evidence of essential unidimensionality – the total variance explained on the general factor was close to an acceptable percentage of 60%. The factor loadings of only items B16, B17 and B20 were not significant on the general factor.

Figure 3 indicates that three factors explained 46.7% of the total variance of Scale C2, and thus, it appears that Scale C2 showed evidence of multidimensionality.

Figure 4 indicates that the general factor accounted for 48% of the variance, and the group factors individually accounted for a greater portion of the variance. Therefore, it is confirmed that Scale C2 showed little evidence of unidimensionality and appeared to be multidimensional.

Upon further investigation, it was found that items C2 27 and C2 B2 had a weak factor loading (0.28 and 0.16, respectively) onto the general factors and the underlying factors. Item C2 42 had an insignificant factor loading onto the general factor, but a significant factor loading onto factor 3 (0.32). As the item measures a factor other than the general factor and could be a contributing factor to the multidimensional appearance of Scale C2, it appears that the above mentioned items need reviewing.

Figure 3 shows that two strong factors explained 56% of the total variance in respect of Scale C6; therefore, this scale showed evidence of multidimensionality.

TABLE 1: Intra-scale correlation matrix, alpha coefficients and descriptive statistics.

Variable	C1	C2	C3	C4	C5	C6	B†	A†
C1	0.82‡	-	-	-	-	-	-	-
C2	0.94	0.70‡	-	-	-	-	-	-
C3	0.61	0.63	0.79‡	-	-	-	-	-
C4	0.39	0.45	0.40	0.84‡	-	-	-	-
C5	0.66	0.69	0.52	0.36	0.82‡	-	-	-
C6	0.69	0.69	0.65	0.41	0.61	0.88‡	-	-
B	0.58	0.64	0.53	0.47	0.59	0.56	0.82‡	-
A	0.41	0.51	0.45	0.22	0.33	0.43	0.37	0.95‡
Standard deviation	5.66	12.08	4.67	7.91	6.49	7.57	12.74	22.76
Mean	24.34	44.72	27.94	29.43	24.94	41.22	44.16	81.2

†, Scale B and A correlation coefficients have been reflected to positives for the purposes of the analyses to enhance interpretation (Likert Scales C1 to C6 are anchored in an opposite direction to Scales B and A); ‡, Values in bold on the diagonal represent the alpha reliability coefficients of the respective scales for the sample group.

FIGURE 3: Findings from Exploratory Factor Analysis on the Subscales of the Experience of Work and Life Circumstances Questionnaire.†

SCALE A — Total % Var: 39.91

Factor 1	Eigen 13.34	% Var 33.34
Item	**Loading**	
A1	0.43	
*A4	0.44	
A5	0.64	
A6	0.75	
A8	0.74	
A7	0.12	
A12	0.60	
A14	0.52	
A15	0.80	
A16	0.73	
A17	0.54	
*A18	0.38	
A21	0.67	
A22	0.22	
A23	0.43	
*A25	0.30	
*A26	0.71	
A27	0.31	
A28	0.66	
A29	0.59	
A38	0.48	
A40	0.44	

Factor 2	Eigen 2.63	% Var 6.57
A2	0.48	
A3	0.55	
*A4	0.31	
A7	0.21	
A9	0.66	
A10	0.58	
A11	0.34	
A13	0.51	
*A18	0.32	
A19	0.36	
A20	0.71	
A22	0.26	
A24	0.49	
*A25	0.41	
*A27	0.37	
A30	0.71	
A31	0.61	
A32	0.58	
A33	0.52	
A34	0.67	
A35	0.36	
A36	0.76	
A37	0.52	
A39	0.69	

SCALE B — Total % Var: 38.32

Factor 1	Eigen 4.61	% Var 28.83
Item	**Loading**	
B8	0.40	
B9	0.78	
B10	0.71	
B11	0.78	
B12	0.31	
B13	0.37	
B17	0.27	
*B19	0.31	
B20	0.28	

Factor 2	Eigen 1.53	% Var 9.49
B14	0.67	
B15	0.62	
B16	0.22	
B18	0.26	
*B19	0.31	
B21	0.50	
B22	0.56	
B23	0.63	

SCALE C1 — Total % Var: 46.29

Factor 1	Eigen 3.70	% Var 46.29
Item	**Loading**	
C1 1	0.60	
C1 18	0.61	
C1 21	0.60	
C1 40	0.55	
C1 41	0.70	
C1 45	0.75	
C1 52	0.74	
B1	0.38	

SCALE C2 — Total % Var: 46.67

Factor 1	Eigen 3.99	% Var 26.59
Item	**Loading**	
C2 4	0.48	
C2 5	0.76	
C2 6	0.79	
C2 8	0.29	

Factor 2	Eigen 1.60	% Var 10.65
C2 3	0.47	
C2 17	0.23	
C2 19	0.72	
C2 20	0.47	
C2 28	0.50	
C2 39	0.63	

Factor 3	Eigen 1.42	% Var 9.43
C2 27	0.29	
C2 42	0.37	
C2 46	0.75	
C2 47	0.70	
B2	0.22	

SCALE C3 — Total % Var: 42.21

Factor 1	Eigen 3.38	% Var 42.21
Item	**Loading**	
C3 9	0.56	
C3 22	0.67	
C3 23	0.70	
C3 24	0.76	
C3 29	0.47	
C3 34	0.30	
C3 37	0.70	
B3	0.42	

SCALE C4 — Total % Var: 44.83

Factor 1	Eigen 4.04	% Var 44.83
Item	**Loading**	
C4 10	0.52	
C4 11	0.70	
C4 12	0.69	
C4 26	0.80	
C4 33	0.59	
C4 43	0.55	
C4 48	0.38	
C4 51	0.76	
B4	0.48	

SCALE C5 — Total % Var: 44.59

Factor 1	Eigen 3.57	% Var 44.59
Item	**Loading**	
C5 14	0.49	
C5 15	0.68	
C5 30	0.69	
C5 31	0.61	
C5 44	0.60	
C5 50	0.70	
C5 53	0.67	
B5	0.35	

SCALE C6 — Total % Var: 56.81

Factor 1	Eigen 5.30	% Var 44.14
Item	**Loading**	
C6 2	0.65	
C6 7	0.64	
C6 25	0.87	
C6 32	0.68	
C6 36	0.54	
C6 38	0.82	
B6	0.76	

Factor 2	Eigen 1.52	% Var 12.67
C6 13	0.52	
C6 16	0.70	
C6 35	0.65	
C6 49	0.78	
B7	0.35	

†, only highest loadings and significant cross-loadings are reported. The complete factor matrix may be obtained from the authors; *, Indicates cross-loading exceeding > 0.30; % Var, Percentage variance explained.

SCALE A

Factor 1		
Item	General	Group
A2	0.50	0.31
A3	0.46	0.34
*A7	0.25	0.14
A9	0.38	0.40
A10	0.33	0.35
A11	0.49	0.23
*A13	0.28	0.31
A19	0.38	0.23
A20	0.46	0.43
A22	0.37	0.18
A24	0.53	0.32
A25	0.56	0.27
A27	0.54	0.25
A30	0.58	0.44
A31	0.49	0.38
A32	0.54	0.37
A33	0.38	0.32
A34	0.35	0.40
A35	0.43	0.24
A36	0.51	0.46
A37	0.58	0.34
A38	0.48	0.18
A39	0.54	0.43
Factor 2		
A1	0.50	-0.28
A4	0.59	-0.29
A5	0.39	-0.38
A6	0.43	-0.45
A8	0.43	-0.44
A12	0.47	-0.37
A14	0.38	-0.32
A15	0.46	-0.48
A16	0.50	-0.44
A17	0.54	-0.34
A18	0.55	-0.26
A21	0.52	-0.41
A23	0.57	-0.28
A25	0.56	-0.21
A26	0.46	-0.43
A28	0.52	-0.41
A29	0.59	-0.37
A28	0.52	-0.31
A40	0.54	-0.29

SCALE B

Factor 1		
Item	General	Group
B8	0.37	0.26
B9	0.58	0.52
B10	0.48	0.47
B11	0.45	0.52
B12	0.41	0.21
B13	0.50	0.25
*B17	0.23	0.18
*B20	0.28	0.18
Factor 2		
B14	0.47	0.45
B15	0.52	0.41
*B16	0.26	0.14
B18	0.33	0.17
B19	0.45	0.20
B21	0.35	0.33
B22	0.54	0.37
B23	0.36	0.42

SCALEC2

Factor 1		
Item	General	Group
C2 4	0.42	0.38
C2 5	0.50	0.59
C2 6	0.33	0.61
C2 8	0.43	0.23
Factor 2		
C2 3	0.51	0.26
C2 17	0.41	0.16
C2 19	0.57	0.39
C2 20	0.53	0.25
C2 28	0.38	0.27
C2 39	0.44	0.34
Factor 3		
*C2 27	0.28	0.25
*C2 42	0.07	0.32
C2 46	0.38	0.66
C2 47	0.36	0.61
*B2	0.16	0.19

SCALE C6

Factor 1		
Item	General	Group
C6 2	0.56	0.41
C6 7	0.55	0.41
C6 25	0.57	0.56
C6 32	0.62	0.43
C6 36	0.56	0.34
C6 38	0.62	0.52
B6	0.47	0.48
Factor 2		
C6 13	0.33	0.33
C6 16	0.51	0.45
C6 35	0.58	0.41
C6 49	0.57	0.50
B7	0.48	0.22

Proportion of variance % explained for bi-factors and factors:

Subscales	General	Group	Factor 1	Factor 2	Factor 3
Scale A	62%	38%	19°%	19%	N/A
Scale B	58%	42%	22%	20%	N/A
Scale C2	48%	52%	21%	10%	21%
Scale C6	60%	40%	26%	14%	N/A

*, indicates loading of < 0.3 onto the general factor.

FIGURE 4: Detailed findings from bi-factor analysis using Schmid–Leiman analysis.

Figure 4 indicates that Scale C6's general factor accounted for 60% of the total variance, therefore, confirming that Scale C6 showed evidence of essential unidimensionality. All the items loaded significantly on the general factor.

The inter-correlation between the scales is depicted in the inter-correlation matrix presented in Table 1. The relevant inter-correlations are presented at the lower end of the diagonal. It was expected that the scales would correlate positively with each other as they were all related to a higher-order construct, namely that of overall work stress level. However, high correlations may resemble weak differentiation between scales as these scales may be measuring the same construct. Weakly differentiated scales may lead to complex

factor structures and significant item cross-loadings when all the WLQ items are included in a factor analysis (Farrell, 2010). Scales C1, C2, C5 and C6 displayed high inter-correlations and may be less well differentiated, especially C1 and C2, which correlated very strongly ($r = 0.94$).

Results of the confirmatory factor analyses

This section reports on the SEM and CFAs for the purposes of investigating the WLQ structural models for Hypothesis 2 and Hypothesis 3.

The standardised factor loadings and regression coefficients for the higher-order structural model for the current sample are presented in Figure 5. The high inter-correlation (0.94) between Scales C1 and C2 shows evidence of multicollinearity at a level that indicates a probability of over 80% that an improper solution will be obtained because of inaccurate estimates of coefficients and standard errors (Grewal, Cote & Baumgartner, 2004). The model fit indices (chi-square [19] = 122.27, $p < 0.05$; Comparative Fit Index (CFI) = 0.91; Tucker-Lewis Index (TLI) = 0.87; Standardised Root Mean Square Residual (SRMR) = 0.07; Root Mean Square Error of Approximation (RMSEA) = 0.16) point to a weak model fit. It would be reasonable to argue that conceptually the content of Scales C1 and C2 showed strong overlap, and therefore, should be treated as such for the purposes of specifying the appropriate SEM model (Byrne, 2008). The model was subsequently respecified and re-estimated with error covariances included between Scales C1 and C2. The model fit indices improved significantly (chi-square [18] = 53.63, $p < 0.05$; CFI = 0.96; TLI = 0.94; SRMR = 0.03; RMSEA = 0.10) to indicate a reasonable model fit. F1 (internal causes of stress) appeared to be well defined, as all the factor loadings were strong, except for Scale C4 (career matters), which displayed a weak loading. Scale B (external causes of stress) appeared to overlap strongly ($R = 0.73$, $p < 0.05$) with F1, which signified that the population sample perceived Scale B and F1 to be related. The non-significant regression coefficient ($R = -0.06$, $p > 0.05$) showed that Scale B had no relationship with Scale A
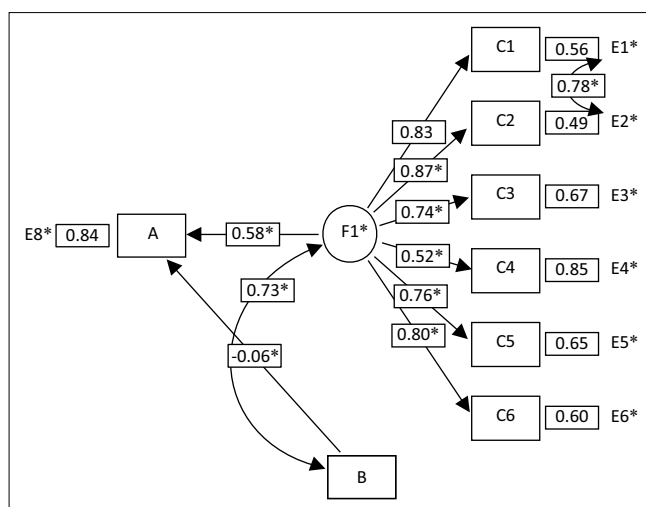
(stress level) and, therefore, there was no evidence that Scale B might have a direct effect on Scale A. The regression coefficient showed that F1 had a statistically significant and moderate to strong correlation ($R = 0.58$, $p < 0.05$) with Scale A, providing some support that F1 may be affecting Scale A. The combined effect of Scale B and Scale F1 explained 29% (R-square = 0.29) of the variance on Scale A and can be considered moderate to strong, which supported a direct relationship with mostly F1 (adverse circumstances at work).
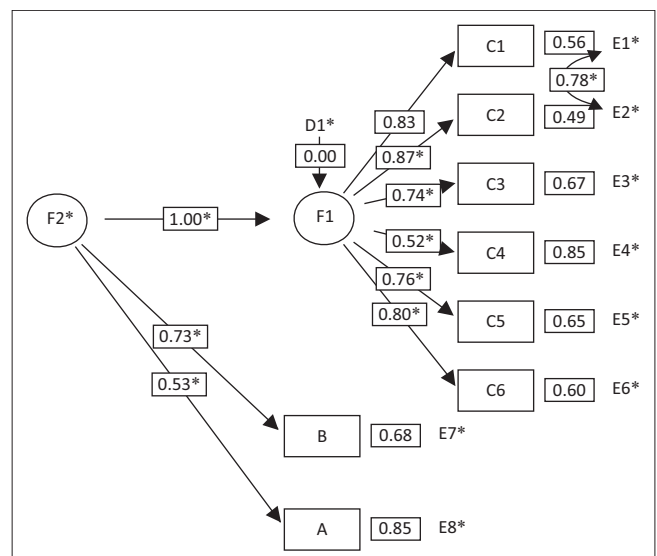
The standardised factor loadings for the higher-order latent variable structural model for the sample are presented in Figure 6. This model was specified and estimated to allow for error covariances between Scale C1 and Scale C2. The model fit indices (chi-square [17] = 45, $p < 0.05$; CFI = 0.97; TLI = 0.95; SRMR = 0.03; RMSEA = 0.09) pointed to a relatively good model fit. F1 (internal causes of stress) was well defined, as all the factor loadings were strong except for Scale C4 (career matters). F1 loaded strongly on the higher-order factor F2 (latent variable: overall work stress level). Scale B (external causes of stress) loaded strongly on F2, but Scale A (level of stress) had a moderate loading on the higher-order factor F2 (overall work stress level).

Discussion

The present study aimed to explore and evaluate the structural validity of WLQ scores in respect of a sample of employees in the automotive assembly plant of an automotive manufacturing company in South Africa.

Outline of results

It is well recognised that sample demographics and contextual factors are likely to have a differential effect on the reliability and validity of test scores (Cizek, 2015; Guion & Highhouse, 2011; Tavakol & Dennick, 2011). The research evidence suggests that the validity of the proposed WLQ

*, indicates free parameters.

FIGURE 5: Higher-order structural model of the Experience of Work and Life Circumstances Questionnaire based on the sample data (N = 217).

*, indicates free parameters.

FIGURE 6: Higher-order latent variable structural model of the Experience of Work and Life Circumstances Questionnaire based on the sample data (N = 217).

models appears to hold for the sample group in question. The majority of the respondents in the sample group were black Africans (54%), had obtained post-matric qualifications (82%), had been employed in the organisation for some time (76% > 2 years) and had an average of one to three dependants (63% < 2 dependants). Most of the subscales appeared to be essentially homogeneous in terms of item meaning, and the validity of score inferences based on the proposed higher-order measurement models developed for the study group, appeared to be supported by the data. The reliability statistics were good to acceptable for all the scales in this study, and it seemed that the sample and the context of use had little effect on the reliability of the WLQ. The reliability coefficients in the current study were consistent with those indicated in the study conducted by Oosthuizen and Koortzen (2009), and Van Zyl and Van der Walt (1991). According to Loewenthal (2001) and Tavakol and Dennick (2011), reliability is closely related to validity; thus, if an assessment is not reliable, it cannot be valid. Yet an assessment can be proven to be reliable but not valid (Tavakol & Dennick, 2011; Tonsing, 2013). The validity of an instrument pertains to the correctness of the inferences that are made on the basis of the test results (Slocum-Gori & Zumbo, 2011). Unidimensionality is an important prerequisite for the correctness of the score inferences that are made. In the present study, the EFA and bi-factor analysis results provided evidence that the scales of the WLQ are at least essentially unidimensional. Therefore, aggregating the individual item scores to produce a scale score would be justified, except in the case of Scale C2 where the level of homogeneity in item meaning was not sufficient to represent an adequately delineated interpretation of the scale and would, therefore, require revision. The high inter-correlations observed between Scale C1 (organisational functioning) and Scale C2 (task characteristics) suggested that these scales lacked deferential validity and might need to be revised. Although the WLQ manual describes the subscales by listing the relevant content that forms part of each scale, a more concise and more carefully delineated definition of the common underlying factors would enhance interpretation (DeVellis, 2012; Slavec & Drnovesek, 2012).

The results obtained from the CFA showed that the structural and higher-order measurement model proposed by Van Zyl and Van der Walt (1991) might be partially valid in respect of the study group, provided that the differential validity problem of the C1 and C2 scales and the negative effect of multicollinearity on model fit were sufficiently addressed. Hence, it was indicated that the data partially supported the WLQ higher-order structural model and the validity of inferences from the modelled interrelationships between the variable scores. However, this may be only one of multiple plausible explanations and, unless this effect is controlled for, SEM users are cautioned not to make unwarranted causal claims (Lei, Wu & Pennsylvania, 2007). The moderate to strong relationship between the experience of stress and work environment stressors as measured by the WLQ is supported in the literature (Robbins & Judge, 2015; Tiyce, Hing, Cairncross & Breen, 2013, p. 146). However, contrary to expectations, the factors external to the work environment appeared not to have had a direct effect on the experience of work stress. Users of the instrument should therefore be cautious about linking the experience of stress directly to sources outside the work environment as this relationship appears to be of a more complex nature. Robbins and Judge (2015) indicate that evidence suggests that people experience the internal work environment (the job) as the single most common source of stress followed by personal finances, health and other factors outside the work environment, but these are notably less prominent. Tiyce *et al.* (2013, p. 146) state that many researchers who conducted studies in various industries concur 'that workplace factors are a fundamental cause of stress for employees'. Edwards, Guppy and Cockerton (2007) posit that research findings support an interaction and inter-correlation between job and non-job sources of stress. These findings most possibly explain the strong positive covariance between external environment stressors and work environment stressors reported in the current study. The exact reasons for these findings and how external and internal sources interact need further exploration. The higher-order structural model of the WLQ was adequately supported by the data obtained in the current study, and, therefore, valid inferences from the proposed latent variable scores could possibly be made. It may therefore be justifiable to aggregate all subscale scores to form a single higher-order, overall stress score consisting of experience of stress, work-related sources of stress and external environment sources of stress.

The WLQ manual provides a priori theoretical propositions relevant to the internal structure of the respective higher-order measurement models presented in this study (Cox, 1978; Cox & McKay, 1978). Moreover, the models suggested by Van Zyl and Van der Walt (1991) have become entrenched in work stress theory (Cartwright & Cooper, 1997; Catellano & Plionis, 2006; Colligan & Higgins, 2005; Mark & Smith, 2008; Nelson & Quick, 2000; Robbins & Judge, 2015). However, grounded theory in support of hypothetical relationships between variables should accompany empirical evidence to ensure valid score inferences (AERA *et al.*, 2014; Cizek, 2015; Slavec & Drnovesek, 2012). A number of studies that support the construct validity of the WLQ were presented earlier in this article (Lamb, 2009; Roos & Van Eeden, 2008; Sardiwalla, 2014; Van Zyl, 2009; Van Zyl & Van der Walt, 1991). Although these studies show that the individual WLQ scales relate significantly to external variables (convergent and discriminant validity), the most important higher-order internal relationships between the experience of stress and the stressors were never tested. It was unexpected that no prior studies investigated the hypotheses related to the higher-order internal structural models proposed by Van Zyl and Van der Walt (1991), and it is not clear why this is the case. Theoretical substantiation of the respective models investigated in this study can be considered to be an important factor as regards the validity of score inferences with respect to the constructs measured.

Practical implications

In the introductory section of this paper, it was argued that the EEA requirements related to the body of evidence required to confirm the validity of a test (for instance, the WLQ) were unknown, which consequently resulted in much uncertainty about the legitimacy of using tests in a specific context. Furthermore, the evidence that published resources provide about tests may be of a generic nature and highly challengeable if generalisability of the evidence across contexts (Cizek, 2015; Guion & Highhouse, 2011) and proof of authenticity (absence of publication bias) cannot be demonstrated (Ferguson & Heene, 2012; McDaniel *et al.*, 2006). A critical stance was taken in this study and balanced reporting of both confirmatory and non-confirmatory findings on the structural validity of the WLQ was done (Guion & Highhouse, 2011). The context-specific research results in this study provide valuable insights and scientific evidence in terms of a continuum of evidentiary support and in terms of the extent to which the WLQ score inferences are valid from an internal structural perspective. The structural validity evidence gathered in this study was very helpful and a notable step towards confirming the generalisability, objectivity and authenticity of the evidence published in the WLQ manual (Van Zyl & Van der Walt, 1991) and elsewhere (Oosthuizen & Koortzen, 2009). In addition, the findings of this study determined that sufficient grounds existed for doing further validity research and for continuing to use the instrument in the company under study. Furthermore, the approach of gathering reliability and validity evidence in context of use is in line with the principles of contemporary validity theory (the 2014 Standards). Considering the uncertain regulatory environment, this approach is more likely to achieve scholarly and expert witness recognition should the WLQ's reliability and validity be challenged legally. However, the process of enquiry into the validity of the WLQ in general may be considered far from complete. It can be deduced that the WLQ shows promise to be used as a diagnostic or selection tool for the specific study group provided that the research continues in pursuit of a more 'unitary' body of evidence. Notwithstanding the findings made in respect of the specific sample group, a number of studies support the construct validity of the WLQ scales through correlations obtained with related measures (Lamb, 2009; Roos & Van Eeden, 2008; Sardiwalla, 2014; Van Zyl, 2009; Van Zyl & Van der Walt, 1991). Validity research involves a continuum of evidentiary support, and findings may vary as plausible alternative internal structures, the effect of context, sample characteristics (e.g. language proficiency levels) and the consequences of the use of a measure are explored.

This paper argues against the notion of relying on HPCSA-certified psychological instruments as being legally compliant, especially because certification is now coupled with EEA legislation, which does not take the evolving nature of the test validation process into account. Such certification and the EEA do not acknowledge that there is still no professional consensus about the concept and meaning of validity, even though this construct has been debated for more than 100 years (Newton & Shaw, 2015). Despite regulatory frameworks or professional best practice testing guidelines that instruct test developers, test reviewers and users to make overall judgements about the validity of a test based on an integration of research evidence related to score inferences and score use, there is no evidence that consensus about a test's overall validity has ever successfully been reached (Cizek, 2015). Validity evidence related to test score meaning is a different issue from that of evidence related to specific test use, as these issues require different validation processes that target different aspects of defensible testing. The paper argues that the responsibility lies with every test user or psychologist to ensure that context-specific research is conducted to support the validity of inferences made on the basis of test scores. Thereafter, a test user should work systematically towards the 'unitary concept' of validity, which requires a collective body of research evidence to ensure more informed judgements about the possible consequences of testing in specific contexts. It is the ethical responsibility of every test user to embark on the process of test validity research for specific uses – indeed, successful litigation in terms of the EEA requirements may depend on such evidence of validity. The HPCSA's training guidelines and frameworks for the qualification and registration of psychometrists (independent practice) (Form 94) and industrial psychologists (Form 214), as well as most master's programmes in industrial psychology in South Africa, make provision for research skills and extensive exposure to psychological testing theory and practice (HPCSA, 2011, 2014b). There are, therefore, sufficient reasons for qualified practitioners in the field of psychological testing in South Africa to embark on validation studies on context-of-test use.

Limitations and recommendations

A limitation of this study was the sampling and sample size constraints related to conducting research in a specific environment with the objective of finding answers to research questions that might apply to a specific population or group, in this case, investigating the structural validity of the WLQ administered to a sample of employees in an automotive assembly plant of an automotive manufacturing company. More comprehensive and rigorous statistical analyses at item level, which usually require larger sample sizes to obtain adequate statistical power, would have enhanced the findings of the study. Such analyses could include an EFA that uses a correlation matrix that includes all WLQ items. Alternatively, the more rigorous CFA, based on a covariance matrix that includes all WLQ items, would allow testing more comprehensive models at item level. However, as a rule of the thumb, such analyses may require a ratio of between 5 and 10 respondents per item, depending on the analysis conducted. The WLQ consists of 116 items and 8 subscales; therefore, effective analysis would need a considerably higher number of respondents than that which was used in this study.

It should be noted that the testing of causal claims with respect to the relationship between sources of stress and experience of work stress requires longitudinal studies and therefore the findings of this study with respect to the higher-order structural model should be considered only as preliminary evidence.

Conclusion

In conclusion, valuable information was obtained in this study that may help to determine whether the relevant company should continue to use the WLQ measuring instrument. The psychometric properties of the WLQ in terms of the unidimensional scale requirement showed adequate reliability and homogeneous item content. However, preliminary evidence suggests that the differential validity of some of the subscales may be questionable and may need further testing. The empirical evidence obtained in this study does support the theory that a direct relationship exists between the experience of stress and work-related environmental sources of stress and that the WLQ subscales may be aggregated to a single overall stress score for the sample group. Although it would be presumptuous to draw any validity conclusions at this stage, it can be said that the context-specific research findings of this study were very helpful in evaluating whether the instrument would be of further use. Further research is required to assess the use of the WLQ in more contexts and on extending the methodology and statistical analysis techniques in pursuit of a 'unitary concept of validity', as defined in the 2014 Standards.

Acknowledgements

Competing interests

Authors' contributions

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Arnold, J., Cooper, C.L., & Robertson, I.T. (1995). *Work psychology: Understanding human behaviour in the workplace*. London: Pitman.

Babbie, E. (2013). *The practice of social research*. (13th edn.). Belmont, CA: Wadsworth.

Baron, H., Henley, S., McGibbon, A., & McCarthy, T. (2002). *Motivation questionnaire manual and user's guide*. Sussex: Saville and Holdsworth.

Bornstein, R.F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment*, 23(2), 532–544. http://dx.doi.org/10.1037/a0022402

Byrne, B.M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, 20, 872–882.

Carter, P. (2011). *IQ and psychometric tests: Assess your personality, aptitude and intelligence*. (2nd edn.). London: Kogan Page.

Cartwright, S., & Cooper, C.L. (1997). *Managing workplace stress*. London: SAGE.

Catellano, C., & Plionis, E. (2006). Comparative analysis of three crisis intervention models applied to law enforcement first responders during 9/11 and hurricane Katrina. *Brief Treatment and Crisis Intervention*, 6(4), 326. http://dx.doi.org/10.1093/brief-treatment/mhl008

Cattell, R.B., Eber, H.W., & Tatsuoka, M.M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16PF)*. Champaign, IL: IPAT.

Cizek, G.J. (2015). Validating test score meaning and defending test score use: Different aims, different methods. *Assessment in Education: Principles, Policy and Practice*, (August), 1–14. http://dx.doi.org/10.1080/0969594X.2015.1063479

Colligan, T.W., & Higgins, E.M. (2005). Workplace stress: Etiology and consequences. *Journal of Workplace Behavioral Health*, 21(2), 89–97. http://dx.doi.org/10.1300/J490v21n02

Costello, A.B., & Osbourne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Validation*, 10(7), 1–9.

Courtney, M.G.R. (2013). The number of factors to retain in EFA: Using the SPSS R-Menu v2.0 to make more judicious estimations. *Practical Assessment, Research and Evaluation*, 18(8), 1–14.

Cox, T. (Ed.). (1978). *Stress*. London: Macmillan.

Cox, T., & McKay, C. (1978). Stress at work. In T. Cox (Ed.), *Stress*, pp. 147–176. Baltimore, MD: University Park Press.

Cox, T., & Mackay, C.J. (1981). A transactional approach to occupational stress. In E.N. Corlett & J. Richardson (Eds.), *Stress, work design and productivity*, pp. 91–114. Chichester: Wiley.

Davies, B., Phil, A., & Warr, P. (2000). *Corporate Culture Questionnaire manual and user's guide*. Sussex: Saville and Holdsworth.

Department of Health. (2006). *Ethical rules of conduct for health professionals registered under the Health Professions Act, 1974*. Government Gazette, pp. 63–68. Pretoria: Government Printing Works.

DeVellis, R.F. (2012) *Scale development: Theory and applications*. (2nd edn.). Thousand Oaks, CA: Sage.

Dumont, E., Kroes, D., Korzilius, H., Didden, R., & Rojahn, J. (2014). Psychometric properties of a Dutch version of Behaviour Problems Inventory (BPI-01). *Research in Development Disabilities*, 35, 603–610. http://dx.doi.org/10.1016/j.ridd.2014.01.003

Edwards, J.A., Guppy, A., & Cockerton, T. (2007). A longitudinal study exploring the relationships between occupational stressors, non-work stressors, and work performance. *Work & Stress*, 21(2), 99–116. http://dx.doi.org/10.1080/02678370701466900

Farrell, A.M. (2010). Insufficient discriminant validity: A comment on Bove, Pervan, Beatty, and Shiu (2009). *Journal of Business Research*, 63(3), 324–327. http://dx.doi.org/10.1016/j.jbusres.2009.05.003

Ferguson, C.J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, 7(6), 555–561. http://dx.doi.org/10.1177/1745691612459059

Fontana, D. (1994). *Problems in practice series: Managing stress*. London: British Psychological Society and Routledge.

Fouché, F.A. & Grobbelaar, P.E. (1983). *Manual for the PHSF Relations Questionnaire*. Pretoria: Human Sciences Research Council.

Foxcroft, C., & Roodt, G. (Eds.). (2013). *Introduction to psychological assessment in the South African context*. (4th edn.). Cape Town: Oxford University Press.

Gaskin, C.J., & Happell, B. (2014). On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. *International Journal of Nursing Studies*, 51, 511–521. http://dx.doi.org/10.1016/j.ijnurstu.2013.10.005

Grewal, R., Cote, J.J.A., & Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Marketing Science*, 23(4), 519–529. http://dx.doi.org/10.1287/mksc.1040.0070

Guion, R.M., & Highhouse, S. (2011). *Essentials of personnel assessment and selection*. New York: Routledge.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139–164. http://dx.doi.org/10.1177/014662168500900204

Hayton, J.C., Allen, D.G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191–205. http://dx.doi.org/10.1177/1094428104263675

Health Professions Council of South Africa. (2010). *Policy on the classification of psychometric measuring devices, instruments, methods and techniques* (Form 208). Retrieved November, 2015, from http://www.hpcsa.co.za/uploads/editor/UserFiles/downloads/psych/F208%20update-%2001%20June%202010-Exco%20(2).pdf

Health Professions Council of South Africa. (2011). *Requirements in respect of internship programmes in industrial psychology* (Form 218). Retrieved November, 2015, from http://www.hpcsa.co.za/Uploads/editor/UserFiles/downloads/psych/psycho_education/form_218.pdf

Health Professions Council of South Africa. (2014a). List of classified and certified psychological tests: Notice 93. *Government Gazette*. (No. 37903, August 15). Retrieved November, 2015, from http://www.hpcsa.co.za/uploads/editor/UserFiles/downloads/psych/BOARD%20NOTICE%2093%20OF%2015%20AUG%202014.pdf

Health Professions Council of South Africa. (2014b). *Requirements in respect of internship programmes in industrial psychology* (Form 94). Retrieved November, 2015, http://http://www.hpcsa.co.za/uploads/editor/UserFiles/downloads/psych/FORM%2094%20-training%20and%20examination%20guidelines%20for%20psychometrists%20-%2002%20june%202014.pdf

Hristov, Z., Tomev, L., Kircheva, D., Daskalova, N., Mihailova, T., Ivanova, V. *et al.* (2003). *Work stress in the context of transition: A case study of education, health and public administration in Bulgaria*. Geneva: International Labour Organisation.

Hu, Q., Schaufeli, W.B., & Taris, T.W. (2011). The Job Demands–Resources model: An analysis of additive and joint effects of demands and resources. *Journal of Vocational Behavior*, *79*, 181–190. http://dx.doi.org/10.1016/j.jvb.2010.12.009

Kane, M.T. (2015). Explicating validity. *Assessment in Education: Principles, Policy and Practice*, August, 1–14. http://dx.doi.org/10.1080/0969594X.2015.1060192

Kubiszyn, T.W., Meyer, G.J., Finn, S.E., Eyde, L.D., Kay, G.G., Moreland, K.L., *et al.* (2000). Empirical support for psychological assessment in clinical health care settings. *Professional Psychology: Research and Practice*, *31*(2), 118–130. http://dx.doi.org/10.1037/0735-7028.31.2.119

Kumar, R. (2011). *Research methodology: A step-by-step guide for beginners.* (3rd edn.). London: SAGE.

Lai, J.-S., Crane, P.K., & Cella, D. (2006). Factor analysis techniques for assessing sufficient unidimensionality of cancer related fatigue. *Quality of Life Research*, *15*, 1179–1190. http://dx.doi.org/10.1007/s11136-006-0060-6

Lamb, S. (2009). *Personality traits and resilience as predictors of job stress and burnout among call centre employees*. Unpublished master's dissertation, University of the Free State, Bloemfontein, South Africa. Retrieved October, 2015, from http://scholar.ufs.ac.za:8080/xmlui/bitstream/handle/11660/1099/LambS-1.pdf?sequence=1

Lazarus, R.S., & Folkman, S. (1984). *Stress, appraisal and coping*. New York: Springer.

Lei, P., Wu, Q., & Pennsylvania, T. (2007). Introduction to structural equation modeling: Issues. *Educational Measurement: Issues and Practice*, *26*, 33–43. http://dx.doi.org/10.1111/j.1745-3992.2007.00099.x

Lemanski, C. (2003). Psychological first aid: After debriefing. *Fire Engineering*, *156*(6), 73–75.

Loewenthal, K.M. (2001). *An introduction to psychological tests and scales.* (2nd edn.). East Sussex: Psychology Press.

Ludik, I.J.G. (1988). *Stres en die verpleegkundige (Stress and the nurse)*. Unpublished masters' thesis, Pretoria: University of Pretoria.

Mark, G.M., & Smith, A.P. (2008). Stress models: A review and suggested new direction. In J. Houdmont & S. Leka (Eds.), *Occupational health psychology. Nottingham: Nottingham University Press* (pp. 111–144). Nottingham: Nottingham University Press.

Matteson, M.T. & Ivansevich, J.M. (1987). *Controlling work stress*. San Francisco, CA: Jossey-Bass Publications.

McDaniel, M.A., Rothstein, H.R., & Whetzel, D.L. (2006). Publication bias: A case study of four test vendors. *Personnel Psychology*, *59*, 927–953. http://dx.doi.org/10.1111/j.1744-6570.2006.00059.x

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (pp. 13–103). Washington, DC: American Council on Education and National Council on Measurement in Education.

Moerdyk, A. (2009). *The principles and practices of psychological assessment*. Pretoria: Van Schaik.

Nelson, D.L., & Quick, J.C. 2000. *Organizational behaviour: Foundations, realities and challenges.* (3rd edn.). Cincinnati, OH: South-Western College Publishing.

Newton, P.E., & Shaw, S.D. (2015). Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assessment in Education: Principles, Policy and Practice*, August, 1–19. http://dx.doi.org/10.1080/0969594X.2015.1037241

O'Connor, B. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments and Computers, 32*(3), 396–402. http://dx.doi.org/10.3758/BF03200807

Oosthuizen, J. (2005). *The relationship between stress and salutogenic functioning amongst employees in a state owned enterprise*. Unpublished master's dissertation, University of the Western Cape, Bellville, South Africa. Retrieved December, 2015, from http://etd.uwc.ac.za/xmlui/handle/11394/224

Oosthuizen, R.M., & Koortzen, P. (2009). Psychometric properties of the Experience of Work and Life Circumstances Questionnaire and the Hopkins Symptom Checklist. *South African Journal of Industrial Psychology, 35*(1), 11–20. http://dx.doi.org/10.4102/sajip.v35i1.492

Pallant, J. (2007). *SPSS survival manual: A step-by-step guide to data analysis using SPSS for Windows (SPSS version 15)*. (3rd edn.). Maidenhead: Open University Press.

Reise, S.P., Moore, T.M., & Haviland, M.G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544–559. http://dx.doi.org/10.1080/00223891.2010.496477

Rempler, D.K., & Van Ryzin, G.G. (2011). *Research methods in practice: Strategies for description and causation.* London: Sage.

Republic of South Africa. (1974). *Health Professions Act, No. 56 of 1974*. Retrieved November, 2015, from www.socdev.gpg.gov.za/Legislation/Documents/HealthProfessionsAct.pdf

Republic of South Africa. (1998). Employment Equity Act, No. 55 of 1998. *Government Gazette*. (No. 19370, October 19, pp. 1–28).

Republic of South Africa. (2013). Employment Equity Amendment Act, No. 47 of 2013. *Government Gazette*. (Vol. 583, No. 37238 of January 16, 2014).

Robbins, S.P., & Judge, T.A. (2015). *Organizational behaviour*. Upper Saddle River, NJ: Pearson.

Roos, W., & Van Eeden, R. (2008). The relationship between employee motivation, job satisfaction and corporate culture. *South African Journal of Industrial Psychology, 34*(1), 54–63. http://dx.doi.org/10.4102/sajip.v34i1.420

Sardiwalla, N. (2014). *The influence of stressors and coping strategies in hospice workers*. Unpublished master's dissertation, Bloemfontein: University of the Free State.

Schmid, J., & Leiman, J.N. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*, 53–61. http://dx.doi.org/10.1007/BF02289209

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. http://dx.doi.org/10.3200/JOER.99.6.323-338

Shepard, L.A. (1993). Evaluating test validity. *Review of Research in Education, 19*, 405–450. http://dx.doi.org/10.2307/1167347

Sireci, S.G. (2015). On the validity of useless tests, *Assessment in Education: Principles, Policy & Practice,* (August), 1–10. http://dx.doi.org/10.1080/0969594X.2015.1072084

Slavec, A., & Drnovesek, M. (2012). A perspective on scale development in entrepreneurship research. *Economic and Business Review, 14*(1), 39–62.

Slocum-Gori, S., & Zumbo, B.D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria for factor analysis. *Social Indicators Research, 102*, 443–461. http://dx.doi.org/10.1007/s11205-010-9682-8

Smith, P.C., & Kendall, L.M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47*, 149–155. http://dx.doi.org/10.1037/h0047060

Smits, I.A., Timmerman, M.E., & Meijer, R.R. (2012). Exploratory Mokken Scale Analysis as a dimensionality assessment tool: Why scalability does not imply unidimensionality. *Applied Psychological Measurement, 36*(6), 516–539. http://dx.doi.org/10.1177/0146621612451050

Straub, D., Boudreau, M.-C., & Gefen, D. (2004). Validation guidelines for IS positivistic research. *Communication of the Association for Information Systems, 13*, 380–427.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53–55. http://dx.doi.org/10.5116/ijme.4dfb.8dfd

Theron, C. (2007). Confessions, scapegoats and flying pigs: Psychometric testing and the law. *South African Journal of Industrial Psychology, 33*(1), 102–117. http://dx.doi.org/10.4102/sajip.v33i1.260

Tiyce, M., Hing, N., Cairncross, G. & Breen, H. (2013). Employee stress and stressors in gambling and hospitality workplaces. *Journal of Human Resources in Hospitality & Tourism, 12*(2), 126–154. http://dx.doi.org/10.1080/15332845.2013.752708

Tonsing, K.N. (2013). Psychometric properties and validation of Nepali version of the Depression Anxiety Stress Scales (DASS-21). *Asian Journal of Psychiatry, 4*, 63–66.

Ullman, J.B. (2006). Structural equation modeling: Reviewing the basics moving forward. *Journal of Personality Assessment, 87*(1), 35–50. http://dx.doi.org/10.1207/s15327752jpa8701_03

Van de Vijver, A.J., & Rothman, S. (2004). Assessment in multicultural groups: The South African case. *South African Journal of Industrial Psychology, 30*(4), 1–7. http://dx.doi.org/10.4102/sajip.v30i4.169

Van Zyl, E.S., & Van der Walt, H.S. (1991). *Manual for the Experience of Work and Life Circumstances Questionnaire (WLQ)*. Pretoria: Human Sciences Research Council.

Van Zyl, P. (2009). *The measurements of levels and causes of work stress of employees in a competitive open medical scheme in the private sector*. Unpublished master's dissertation, Pretoria: University of Pretoria,. Retrieved November, 2015, from http://upetd.up.ac.za/thesis/available/etd-03152012-170001/>C12/4/69/gm

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., Van der Maas, H.L.J., & Kievit, R.A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7*, 627–633. http://dx.doi.org/10.1177/1745691612463078

Weng, L.-J., & Cheng, C.-P. (2005). Parallel analysis with unidimensional binary data. *Educational and Psychological Measurement, 65*(5), 697–716. http://dx.doi.org/10.1177/0013164404273941

Wolff, H.-G., & Preising, K. (2005). Exploring item and higher order factor structure with the Schmid-Leiman solution: Syntax codes for SPSS and SAS. *Behavior Research Methods, 37*(1), 48–58. http://dx.doi.org/10.3758/BF03206397