A PSYCHOMETRIC ANALYSIS OF THE UTRECHT WORK ENGAGEMENT SCALE IN THE SOUTH AFRICAN POLICE SERVICE

K. STORM
S. ROTHMANN
WorkWell: Research Unit for People, Policy and Performance,
Faculty of Economic and Management Sciences
PU for CHE

ABSTRACT

The objectives of this research were to validate the Utrecht Work Engagement Scale (UWES) for the South African Police Service (SAPS) and to determine its construct equivalence and bias in different race groups. A cross-sectional survey design was used. Stratified random samples (*N* = 2396) were taken of police members of nine provinces in South Africa. The UWES and a biographical questionnaire were administered. Structural equation modelling confirmed a 3-factor model of work engagement, consisting of Vigour, Dedication and Absorption. These three factors have acceptable internal consistencies. Exploratory factor analysis with target rotations showed equivalence of the three factors for different race groups in the SAPS. No evidence was found for uniform or non-uniform bias of the items of the UWES for different race groups.

OPSOMMING

Die doelstellings van hierdie navorsing was om die Utrecht-werksbegeesteringskaal (UWES) te valideer vir die Suid-Afrikaanse Polisiediens (SAPD) en die konstrukekwivalensie daarvan vir verskillende rassegroepe te bepaal. 'n Dwarssnee opname-ontwerp is gebruik. Gestratifiseerde ewekansige steekproewe (*N* = 2396) is van polisielede uit nege provinsies geneem. Die UWES en 'n biografiese vraelys is afgeneem. Strukturele vergelykingsmodellering het 'n 3-faktormodel, bestaande uit Energie, Toewyding en Absorpsie, aangetoon. Hierdie drie faktore het aanvaarbare interne konsekwentheid getoon. Eksploratiewe faktoranalise met teikenrotasies het konstrukekwivalensie vir die drie faktore vir verskillende rassegroepe in die SAPD getoon. Bewyse is nie gevind vir uniforme of nie-uniforme sydigheid van die items van die UWES vir verskillende rassegroepe nie.

Before the second World War, the field of Psychology had three distinct missions: 1) curing mental illness; 2) making the lives of all people better, more productive and fulfilling; and 3) identifying and nurturing exceptional talent or genius (Seligman & Csikszentmihalyi, 2000). However, after World War II, Psychology became a science that focused largely on healing. Despite its name, research in the Occupational Health Psychology is dramatically weighted on the side of ill-health and unwell-being instead of health and well-being at work. Even the meaning of basic terms are negatively biased – typical usage equates health with the *absence of illness* rather than the *presence of wellness*. Furthermore, the number of articles researching negative aspects outnumbers those dealing with positive aspects by a ratio of 17 to 1 (Diener, Suh, Lucas & Smith, 1999). Myers (2000) arrived at a more favourable ratio of 14 to 1. According to Schaufeli and Bakker (2001), the *Journal of Occupational Health Psychology* published only 6% articles that examined positive aspects of health and well-being. The remaining 94% dealt among others with Repetitive Strain Injury, burnout, violence, discrimination, alcoholism, Post-Traumatic Stress Syndrome, conflict, sleep disorders and negative affect. This almost exclusive attention to pathology neglects the fulfilled individual and the thriving community.

However, there seems to be a more general trend emerging with the recent introduction of the so-called "positive psychology", which shed a new light on the object of Occupational Health Psychology. This positive psychology is seen as an alternative to the predominant focus on pathology and deficits. The aim of this new paradigm is to begin to catalyse a change in the focus of psychology from preoccupation only with repairing the worst things in life to also building positive qualities (Seligman & Csikszentmihalyi, 2000). So the focus of this paradigm is on human strengths and optimal functioning rather than on weaknesses and malfunctioning.

A movement in the direction of positive psychology is also evident in South Africa. The work of Strümpfer (1995, 2002a)

focuses on the fortigenic paradigm, which is different from the dominant pathogenic orientations. This paradigm implies a shift from a psychology of the sick and dysfunctional to a positive psychology with respect to challenges and opportunities of people in the work place. Thus, the fortigenic paradigm focuses on the origins of strength. Recent work of Strümpfer (2002b) also focused on the fortigenic paradigm and its relation to burnout. He considered psychological constructs that could help understand alternatives to burnout, as well as helping people to move in the general direction of health. Wissing and Van Eeden (2002) also focused on positive psychology in their study of psychological well-being. Their study focused on achieving greater empirical clarification of the nature of psychological well-being by investigating the nature of psychological well-being from a fortigenic perspective.

Viewed from this "positive" perspective, it is not surprising that burnout research seems to shift towards its opposite: work engagement. Work engagement is being defined as an energetic state in which the employee is dedicated to excellent performance at work and is confident of his or her effectiveness (Schutte, Toppinen, Kalimo & Schaufeli, 2000). As mentioned above, Occupational Health Psychology focused on the negative effects of work that contributed to burnout. But the question asked is why certain workers can accomplish large amounts of work with enthusiasm and pleasure, without becoming sick or being burned out. Research on work engagement could answer this question.

The concept of engagement is also applicable to police work. Two decades of research in the police stress literature has left little information known about the extent to which policing is stressful. This resulted in relatively little being known about the quality of life among police officers (Hart, Wearing & Headey, 1995). In their attempts to identify the sources of police stress, researchers have focused almost exclusively on the negative aspects of policing (e.g. Band & Manuelle, 1987; Greller, Parsons, & Mitchell, 1992). This resulted in an overall focus of psychological stress in policing and thus an absence of well-being. It is therefore also necessary to study police work in a

positive way. This could be done by focusing on the concept of work engagement or the different levels of engagement experienced by police officers.

It is important to use a valid and reliable instrument when work engagement is measured. Schaufeli, Salanova, González-Romá and Bakker (2002) developed the Utrecht Work Engagement Scale (UWES) and found acceptable reliability for it. Two recent studies using confirmatory factor analysis demonstrated the factorial validity of the UWES (Schaufeli et al., 2002; Schaufeli, Martinez, Pinto, Salanova, & Bakker in press). However, the UWES has not yet been standardised for police officers in the SAPS and no information is available on its reliability and validity (see Rothmann, 2002). This makes it difficult to assess the levels of engagement of police officers and to compare the levels of engagement in various demographic groups, as well as to place research results in context. Therefore, it is necessary to validate the UWES for police officers in the SAPS.

South Africa is a multicultural society and the SAPS employs individuals of diverse cultural backgrounds. Within the South African context it cannot be taken for granted that scores obtained in one culture can be compared across cultural groups. Before comparing scores across cultural groups, equivalence and bias should be tested (Van de Vijver & Leung, 1997). Without a test of equivalence and bias it is impossible to know to what extent scores or constructs underlying an instrument can be compared across cultures.

The objectives of this study were to determine the construct validity and internal consistency of the UWES and to test its construct equivalence and bias for different race groups in the SAPS.

Work engagement
Research on the work engagement concept has taken two different but related paths. Maslach and Leiter (1997) rephrased burnout as an erosion of engagement with the job. Work that started out as important, meaningful and challenging, becomes unpleasant, unfulfilling and meaningless. In the view of these authors, work engagement is characterised by energy, involvement and efficacy, which are considered the direct opposites of the three burnout dimensions, namely exhaustion, cynicism and lack of professional efficacy respectively. Therefore, they also assess work engagement by the opposite pattern of scores on the three Maslach Burnout Inventory (MBI) dimensions – low scores on exhaustion and cynicism, and high scores on efficacy are indicative for engagement.

Schaufeli and his colleagues partly agree with Maslach and Leiter's (1997) description, but take a different perspective and define and operationalise work engagement in its own right. Schaufeli et al. (2002) consider burnout and work engagement to be opposite concepts that should be measured independently with different instruments. Furthermore, burnout and engagement may be considered two prototypes of employee well-being that are part of a more comprehensive taxonomy constituted by the two independent dimensions of pleasure and activation (Watson & Tellegen, 1985). Activation range from exhaustion to vigour, while identification range from cynicism to dedication. According to this framework, burnout is characterised by a combination of exhaustion (low activation) and cynicism (low identification), whereas engagement is characterised by vigour (high activation) and dedication (high identification).

Based on this theoretical reasoning and after in-depth interviews were carried out with engaged employees, Schaufeli and his colleagues have defined engagement as a positive, fulfilling, work-related state of mind that is characterised by vigour, dedication, and absorption. Rather than a momentary and specific state, engagement refers to a more persistent and pervasive affective-cognitive state that is not focused on any particular object, event, individual or behaviour. Work engagement consists of the following dimensions (Schaufeli et al., 2002):

- **Vigour** is characterised by high levels of energy and mental resilience while working, the willingness to invest effort in one's work, not being easily fatigued, and persistence even in the face of difficulties.
- **Dedication** is characterised by deriving a sense of significance from one's work, by feeling enthusiastic and proud about one's job, and by feeling inspired and challenged by it.
- **Absorption** is characterised by being totally and happily immersed in one's work and having difficulties detaching oneself from it. Time passes quickly and one forgets everything else that is around.

Work engagement is also distinct from other established constructs in organisational psychology, such as organisational commitment, job satisfaction or job involvement (Maslach, Schaufeli & Leiter, 2001). *Organisational commitment* refers to an employee's allegiance to the organisation that provides employment. The focus is on the organisation, where engagement focuses on the work itself. *Job satisfaction* is the extent to which work is a source of need fulfilment and contentment, or a means of freeing employees from hassles or things causing dissatisfaction; it does not encompass the person's relationship with the work itself. *Job involvement* is similar to the involvement aspect of engagement with work, but does not include the energy and effectiveness dimensions (Maslach et al., 2001). Lastly, engagement (especially absorption) comes close to what has been called "*flow*", a term used by Csikszentmihalyi (1990) that represents a state of optimal experience that is characterised by focused attention, a clear mind and body unison, effortless concentration, complete control, loss of self-consciousness, distortion of time and intrinsic enjoyment. However, flow is a more complex concept that includes many aspects and refers to rather particular, short-term "peak" experiences instead of a more pervasive and persistent state of mind, as is the case with engagement (Schaufeli et al., 1999).

The measurement of work engagement
Regarding the measurement of work engagement, Schaufeli et al. (2002) disagree with Maslach and Leiter (1997), who stated that engagement is adequately measured by the opposite profile of MBI scores. Schaufeli et al. (2002) argue that, by using the MBI for measuring work engagement, it is impossible to study its relationship with burnout empirically since both concepts are considered to be opposite poles of a continuum that is covered by one single instrument (the MBI). Although they agree that work engagement is the positive antithesis of burnout, they acknowledge that the measurement and the structures of both concepts differ.

Schaufeli et al. (2002) developed a self-report questionnaire to assess work engagement (the Utrecht Work Engagement Scale – UWES), which includes items such as: "I am bursting with energy in my work" (vigour); "My job inspires me" (dedication); "I feel happy when I'm engrossed in my work" (absorption).

Regarding the psychometric qualities of the UWES, preliminary results show that the three engagement scales have sufficient internal consistencies (Schaufeli et al., 2002; in press). For samples one (314 undergraduate students) and two (619 employees) respectively, the Cronbach α's were as follows: Vigour (9 items), $\alpha = 0,68$ and $0,80$; Dedication (8 items), $\alpha = 0,91$ (both samples); Absorption (7 items), $\alpha = 0,73$ and $0,75$. In the student's sample, the value of α could be improved for Vigour when three items were eliminated ($\alpha = 0,78$). The three scales are moderately to strongly related (mean $r = 0,63$ in Sample 1 and mean $r = 0,70$ in Sample 2). Also, the fit of the hypothesised three-factor model to the data is superior to a one-factor solution (Maslach et al., 2001; Schaufeli et al., 2002).

When work engagement measures are applied to different cultural groups (especially when engagement levels for different cultural groups are compared), issues of measurement bias and equivalence become important (Van de Vijver & Tanzer, 1997). According to Van de Vijver and Leung (1997), equivalence and bias of measuring instruments should be computed in each study that takes place in a multicultural or cross-cultural context.

Van de Vijver and Leung (1997) made a hierarchical distinction of three types of equivalence. The first type, namely construct equivalence, indicates the extent to which the same construct is measured across all cultural groups studied. When an instrument measures different constructs in different cultures, i.e. when cultural equivalence exists, no comparison can be made. The same construct is measured in the case of construct equivalence (also labelled structural equivalence). The second type of equivalence is called measurement unit equivalence and can be obtained when two metric measures have the same measurement unit but have different origins. The third type of equivalence is called scalar equivalence and can be obtained when two metric measures have the same measurement unit and the same origin. Equivalence cannot be assumed but should be established and reported in each study (Van de Vijver & Leung, 1997). Construct equivalence is the most frequently studied type of equivalence. No studies of construct equivalence of the UWES in South Africa were found.

If unacceptable construct equivalence is found, item bias should be computed. An item is an unbiased measure of a theoretical construct, for example engagement, if persons from different cultural groups who are equally engaged have the same average score on the item (Van de Vijver & Leung, 1997). Persons with an equal standing on the theoretical construct underlying the instrument should have the same expected score on the item, irrespective of group membership. The definition of bias does not stipulate that the averages of cultural groups should be identical, but only that these averages should be identical across cultural groups for persons who are equally engaged.

Item bias can be produced by sources such as incidental differences in appropriateness of the item content and inadequate item formulation. Bias will lower the equivalence of a measuring instrument. Two types of item bias are distinguished, namely uniform bias and non-uniform bias (Van de Vijver & Leung, 1997). Uniform bias refers to influences of bias on scores that are more or less the same for all score levels. Non-uniform bias refers to influences that are not identical for all score levels.

The above discussion leads to the following hypotheses:
H1: Work engagement, as measured by the UWES, is a three dimensional construct and the UWES shows high internal consistency.
H2: Work engagement is an equivalent and unbiased construct for White, Black, Coloured and Indian police members.

METHOD

Research design
A survey design was used to reach the research objectives. The specific design is the cross-sectional design, where a sample is drawn from a population at one time (Shaughnessy & Zechmeister, 1997).

Study population
Random samples (N = 2396) were taken from police stations in the Limpopo Province, Gauteng, Free State, Mpumalanga, Northern Cape, Western Cape, Eastern Cape, KwaZulu-Natal and North-West Province. Stations were divided into small (fewer than 25 staff members), medium (25 – 100 staff members) and large (more than 100 staff members) stations. All police members at randomly identified small and medium stations in each of the provinces were asked to complete the questionnaire. In the large stations stratified random samples were taken according to sex and race. Table 1 presents some of the characteristics of the participants.

TABLE 1
CHARACTERISTICS OF THE PARTICIPANTS

Item	Category	Percentage
Race	White	41,23
	Black	40,97
	Coloured	13,38
	Indian	3,64
Rank	Constable	7,54
	Sergeant	19,16
	Captain	23,33
	Inspector	43,73
	Senior Superintendent	3,06
	Other	3,20
Province	North West Province	15,86
	Gauteng	9,77
	Mpumalanga	7,30
	Limpopo Province	8,01
	KwaZulu-Natal	10,73
	Free State	13,86
	Eastern Cape	11,64
	Northern Cape	8,89
	Western Cape	13,94
Size of station	Small	31,45
	Medium	39,05
	Large	29,51
Education	Grade 10	11,01
	Grade 11	5,18
	Grade 12	55,98
	Technical college diploma	2,86
	Technikon diploma	20,70
	University degree	2,16
	Postgraduate degree	2,11
Gender	Male	77,08
	Female	22,92
Marital status	Single	19,56
	Married	53,06
	Divorced	23,97
	Separated	2,11
	Remarried	1,30

The sample was mostly male (77,08%), married, and had a high school education. The mean age of participants was 34,53 years, while the mean length of work experience was 12,96 years.

Measuring battery
The *Utrecht Work Engagement Scale* (UWES) (Schaufeli et al., 2002) was used to measure the levels of engagement. Although work engagement is conceptually seen as the positive antithesis of burnout, it is operationalised in its own right. Work engagement is a concept that includes three dimensions: vigour, dedication and absorption. Engaged workers are characterised by high levels of vigour and dedication, and they are immersed in their jobs. It is an (empirical) question whether engagement and burnout are endpoints of the same continuum or if they are two distinct but related concepts. The UWES is scored on a seven-point frequency rating scale,

varying from 0 ("never") to 6 ("always"). The alpha coefficients for the three sub-scales varied between 0,68 and 0,91. The alpha coefficient could be improved (α varies between 0,78 and 0,89 for the three sub-scales) by eliminating a few items without substantially decreasing the scale's internal consistency.

Statistical analysis
The statistical analysis was carried out by means of the SAS program (SAS Institute, 2000). Cronbach alpha coefficients and inter-item correlation coefficients were used to assess the reliability of the UWES (Clark & Watson, 1995). Descriptive statistics (e.g. means, standard deviations, skewness and kurtosis) were used to analyse the data.

Construct (structural) equivalence was used to compare the factor structures of the UWES for the different cultural groups included in the study. Exploratory factor analysis and target (Procrustean) rotation were used to determine construct equivalence (Van de Vijver & Leung, 1997). According to Van de Vijver and Leung (1997), it is not acceptable to conduct factor analyses for different cultural groups to address the similarity of factor-analytic solutions because the spatial orientation of factors in factor analysis is arbitrary. Rather, prior to an evaluation of the agreement of factors in different cultural groups, the matrices of loadings should be rotated with regard to each other (i.e., target rotations should be carried out). The factor loadings of separate groups are rotated either to one target group or to a joint common matrix of factor loadings. After target rotation had been carried out, factorial agreement was estimated using Tucker's coefficient of agreement (Tucker's phi). This coefficient is insensitive to multiplications of the factor loadings, but is sensitive to a constant added to all loadings of a factor. The following formula is used to compute Tucker's phi (Van de Vijver & Leung, 1997):

$$p_{xy} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 y_i^2}}$$

This index does not have a known sampling distribution hence it is impossible to establish confidence intervals. Values higher than 0,95 are seen as evidence of factorial similarity, whereas values lower than 0,85 are taken to point to non-negligible incongruities (Van de Vijver & Leung, 1997). This index is sufficiently accurate to examine factorial similarity at a global level. However, if construct equivalence is not acceptable, bias analyses should be carried out to detect inappropriate items.

An extension of Cleary and Hilton's (1968) use of analysis of variance was applied to identify item bias (Van de Vijver & Leung, 1997). Bias was examined for each item separately. The item score was the dependent variable, while race groups (four levels) and score levels were the independent variables. Score groups were composed on the basis of the total score on the UWES. A total of ten score levels were obtained by making use of percentiles identified through SAS UNIVARIATE. This made it possible to use score groups with at least 50 persons each. Two effects were tested through analysis of variance, namely the main effect of culture and the interaction of score level and culture. When both the main effect of culture and the interaction of score level and culture are non-significant, the item is taken to be unbiased.

Structural equation modelling (SEM) methods as implemented by AMOS (Arbuckle, 1997) were used to test the factorial model for the UWES, using the maximum likelihood method. Before performing SEM, the frequency distributions of the UWES were checked for normality and multivariate outliers were removed. However, the data did not have a multivariate normal distribution, one of the critically important assumptions associated with SEM. One approach to handling the presence of multivariate non-normal data is to use a procedure known as "the bootstrap" (West, Finch, & Curran, 1995; Yung & Bentler, 1996; Zhu, 1997).

Bootstrapping serves as a resampling procedure by which the original sample is considered to represent the population. Multiple subsamples of the same size as the parent sample are then drawn randomly, *with replacement*, from this population and provide the data for empirical investigation of the variability of parameter estimates and indexes of fit (Byrne, 2001). The underlying concept of the bootstrap technique is that it enables one to create multiple subsamples from an original database in order to examine parameter distributions relative to each of these spawned samples, thereby reporting values with a greater degree of accuracy (Byrne, 2001).

Hypothesised relationships are tested empirically for goodness of fit with the sample data. The χ^2 statistic and several other goodness-of-fit indexes summarise the degree of correspondence between the implied and observed covariance matrixes. Jöreskog and Sörbom (1993) suggest that the χ^2 value may be considered more appropriately as a badness-of-fit, rather than as a goodness-of-fit measure in the sense that a small χ^2 value is indicative of good fit. However, because the χ^2 statistic equals $(N-1)F_{min}$, this value tends to be substantial when the model does not hold and the sample size is large (Byrne, 2001). A large χ^2 relative to the degrees of freedom indicates a need to modify the model to better fit the data. Researchers have addressed the χ^2 limitations by developing goodness-of-fit indexes that take a more pragmatic approach to the evaluation process. One of the first fit statistics to address this problem was the χ^2/degrees of freedom ratio (CMIN/DF) (Wheaton, Muthén, Alwin & Summers, 1977). These criteria, commonly referred to as "subjective" or "practical" indexes of fit are typically used as adjuncts to the χ^2 statistic.

The Goodness of Fit Index (GFI) indicates the relative amount of the variances/co-variances in the sample predicted by the estimates of the population. It usually varies between 0 and 1, and a result of 0,90 or above indicates a good model fit. In addition, the Adjusted Goodness-of-Fit Index (AGFI) is given. The AGFI is a measure of the relative amount of variance accounted for by the model, corrected for the degrees of freedom in the model relative to the number of variables. Although both indexes range from zero to 1,00, the distribution of the AGFI is unknown, therefore no statistical test or critical value is available (Jöreskog & Sörbom, 1986). The parsimony goodness-of-fit index (PGFI) addresses the issue of parsimony in SEM (Mulaik et al., 1989). The PGFI takes into account the complexity (i.e., number of estimated parameters) of the hypothesised model in the assessment of overall model fit and provides a more realistic evaluation of the hypothesised model. Mulaik et al. (1989) suggested that indexes in the 0,90's accompanied by PGFI's in the 0,50's are not unexpected, however, values > 0,80 are considered to be more appropriate (Byrne, 2001).

The Normed Fit Index (NFI) is used to assess global model fit. The NFI represents the point at which the model being evaluated falls on a scale running from a null model to perfect fit. This index is normed to fall on a 0 to 1 continuum. Marsh, Balla and Hau (1996) suggest that this index is relatively insensitive to sample sizes. The Comparative Fit Index (CFI) represents the class of incremental fit indexes in that it is derived from the comparison of a restricted model (i.e., one in which structure is imposed on the data) with that of an independence (or null) model (i.e., one in which all correlations among variables are zero) in the determination of goodness-of-fit. The Tucker-Lewis Index (TLI) (Tucker & Lewis, 1973), which is a relative measure of covariation explained by the model that is specifically developed to assess factor models. For these fit indexes (NFI, CFI and TLI), it is more or less generally accepted that a value of less

than 0,90 indicates that the fit of the model can be improved (Hoyle, 1995), although a revised cut-off value close to 0,95 has recently been advised (Hu & Bentler, 1999).

To overcome the problem of sample size, Browne and Cudeck (1993) suggested using the Root Mean Square Error of Approximation (RMSEA) and the 90% confidence interval of the RMSEA. The RMSEA estimates the overall amount of error; it is a function of the fitting function value relative to the degrees of freedom. The RMSEA point estimate should be 0,05 or less and the upper limit of the confidence interval should not exceed 0,08. Hu and Bentler (1999) suggested a value of 0,06 to be indicative of good fit between the hypothesised model and the observed data. MacCallum, Browne, and Sugawara (1996) recently elaborated on these cut-off points and noted that RMSEA values ranging from 0,08 to 0,10 indicate mediocre fit, and those greater than 0,10 indicate poor fit.

RESULTS

Structural equation modelling (SEM) methods as implemented by AMOS (Arbuckle, 1997) were used to test two factorial models for the UWES, a three-factor as well as a one-factor model of work engagement. It was assumed that the χ^2 goodness-of-fit statistics are not likely to be inflated if the skewness and kurtosis for individual items do not exceed the critical values of 2,0 and 7,0, respectively (West et al., 1995). Data-analyses proceeded as follows: First, a quick overview of each model fit was done by looking at the overall χ^2 value, together with its degrees of freedom and probability value. Global assessments of model fit were based on several goodness-of-fit statistics (GFI, AGFI, PGFI, NFI, TLI, CFI and RMSEA). Secondly, given findings of an ill-fitting initially hypothesised model, analyses proceeded in an exploratory mode using both EFA and CFA. Possible misspecifications as suggested by the so-called modification indexes and standardised residuals values were looked for and eventually a revised, re-specified model was fitted to the data.

Hypothesised three-factor model
The full hypothesised 3-factor model consisting of all 17 items was tested initially. Table 2 presents fit statistics for the test of the original model.

TABLE 2
GOODNESS-OF-FIT STATISTICS FOR THE HYPOTHESISED 3-FACTOR UWES MODEL

Model	χ^2	χ^2/df	GFI	AGFI	PGFI	NFI	TLI	CFI	RMSEA
Default model	1978,79	17,06	0,90	0,87	0,68	0,92	0,91	0,92	0,08

The SEM analyses showed that the 3-factor solution was not admissible. Furthermore, the statistically significant χ^2 value of 1978,79 (df = 116; p = 0,00) revealed a poor overall fit of the originally hypothesised 3-factor UWES model. However, both the sensitivity of the likelihood ratio test to sample size and its basis on the central χ^2 distribution, which assumes that the model fits the population perfectly, have been reported to lead to problems of fit. Jöreskog and Sörbom (1993) pointed out that the use of χ^2 is based on the assumption that the model holds exactly in the population, which is a stringent assumption. A consequence of this assumption is that models that hold approximately in the population will be rejected in a large sample. Furthermore, the hypothesised model (Model 1) was also not that good from a practical perspective. The PGFI value of lower than 0,80, NFI, TLI and CFI values of lower than 0,95 and the RMSEA value of higher than 0,05 are indicative of failure to confirm the hypothesised model. Thus,

it is apparent that some modification in specification is needed in order to determine a model that better represents the sample data.

To pinpoint possible areas of misfit, modification indexes were examined. Furthermore, standardised residuals values were examined. Standardised residuals are fitted residuals divided by their asymptotically (large sample) standard errors (Jöreskog & Sörborn, 1988). In essence, they represent estimates of the number of standard deviations the observed residuals are from the zero residuals that would exist if model fit were perfect (Byrne, 2001). Values > 2,58 are considered to be large (Jöreskog & Sörborn, 1988).

Post hoc analyses
Given rejection of the initially postulated 3-factor model, the focus shifted from model test to model development (exploratory factor analysis). Considering the high standardised residuals of two items, it was decided to re-specify the model with Item 4 and Item 14 deleted. Modification indexes (MI) were also considered to pinpoint areas of misspecification in the model. The constrained parameters exhibiting the highest degree of misfit lay in the error covariance matrix and represent a correlated error between Item 8 and Item 9 (MI = 117,10), as well as between Item 15 and Item 16 (MI = 125,23). Compared with MI values for all other error covariance parameters, these values are exceptionally high and clearly in need of re-specification. Based on the modification indexes and on theoretical considerations, Model 1 was re-specified with these parameters freely estimated. Errors of two item pairs (i.e. VI8-AB9; VI15-AB16) were allowed to correlate. All subsequent analyses are now based on the 15-item revision, which is labelled here as Model 2. The fit statistics are presented in Table 3.

TABLE 3
GOODNESS-OF-FIT STATISTICS FOR MODEL 2 OF THE 3-FACTOR STRUCTURE

Model	χ^2	χ^2/df	GFI	AGFI	PGFI	NFI	TLI	CFI	RMSEA
Default model	1130,28	13,30	0,94	0,91	0,66	0,94	0,93	0,95	0,07

The fit statistics in Table 3 indicate a better fit for the re-specified model. Although the χ^2 value (df = 85; p = 0,00) is still high, it is considerably lower than those in Model 1. All the other fit statistics indicate acceptable fit of the measurement model to the data, although the RMSEA value is still a bit high. Since this model fit was satisfactory and the results agreed with the theoretical assumptions underlying the structure of the UWES according to Schaufeli et al. (2002), no further modifications of the model were deemed necessary. The correlations between the three engagement dimensions were high. Vigour and Dedication show the highest correlation of 0,97, followed by Vigour and Absorption with a correlation of 0,96, and Dedication and Absorption with a correlation of 0,90. The re-specified three-factor model is illustrated in Figure 1.

Following Schaufeli et al. (in press), a unidimensional model was assessed as well. This model assumes that all 17 UWES items load on one single factor. Table 4 presents fit statistics for the test of the original one-factor model.

TABLE 4
GOODNESS-OF-FIT STAISTICS FOR THE HYPOTHESISED 1-FACTOR UWES MODEL

Model	χ^2	χ^2/df	GFI	AGFI	PGFI	NFI	TLI	CFI	RMSEA
Default model	2250,37	18,91	0,87	0,85	0,68	0,90	0,90	0,91	0,09

The statistically significant χ^2 value of 2250,37 (df = 119; p = 0,00) revealed a poor overall fit of the originally hypothesised UWES model. Again, this could be as a result of the large sample size (Jöreskog & Sörbom, 1993). Furthermore, the PGFI value of lower than 0,80, NFI, TLI and CFI values of lower than 0,95 and a high RMSEA value of 0,09 are indicative of failure to confirm the hypothesised model. Therefore, modification indexes as well as standardised residuals were examined.

Post hoc analyses
Based on the high standardised residuals, it was decided to re-specify the 1-factor model with four items deleted (Items 3, 11, 15 and 16). After reviewing the modification indexes, it was decided that the model fit might be further improved by allowing error terms to correlate between Item 4 and Item 5 and between Item 8 and Item 9. In summary, this model was based on 13 of the original 17 items and included correlated errors. In reviewing results bearing on the analysis of this model, Table 5 summarises the goodness-of-fit statistics.

TABLE 5
GOODNESS-OF-FIT STATISTICS FOR MODEL 2 OF THE 1-FACTOR STRUCTURE

Model	χ^2	χ^2/df	GFI	AGFI	PGFI	NFI	TLI	CFI	RMSEA
Default model	777,52	12,34	0,95	0,93	0,66	0,96	0,95	0,96	0,06

The fit statistics in Table 5 indicate a good fit for the re-specified model. Although the χ^2 value (df = 63; p = 0,00) is still high, it is considerably lower than those in Model 1. All the other fit statistics indicate excellent fit of the measurement model to the data. Since this model fit was satisfactory, no further modifications of the model were considered.

The descriptive statistics, alpha coefficients and inter-item correlations of the three factors of the UWES are given in Table 6.

TABLE 6
DESCRIPTIVE STATISTICS, ALPHA COEFFICIENTS AND INTER-ITEM CORRELATIONS OF THE UWES

Item	Mean	SD	Skewness	Kurtosis	r(Mean)	α
Vigour	21,04	6,27	-0,69	0,16	0,42	0,78
Dedication	22,79	6,78	-0,98	0,40	0,62	0,89
Absorption	20,71	6,37	-0,62	0,01	0,41	0,78

The Cronbach alpha coefficients of the scales are considered to be acceptable compared to the guideline of α < 0,70 (Nunnally & Bernstein, 1994). Furthermore, the inter-item correlations are considered acceptable compared to the guideline of 0,15 < r < 0,50 (Clark & Watson, 1995). It appears that the scales have acceptable levels of internal consistency.

Although it seems as if the 1-factor model fitted the data better than the 3-factor model, this is based only on slightly better goodness-of fit indices, and after four items were deleted. Therefore, these results provide support for Hypotheses 1.

Next, exploratory factor analysis and target (Procrustean) rotation were used to determine the construct equivalence of the UWES. The factor loadings of race groups were rotated to one target group. Factorial agreement was estimated using Tucker's coefficient of agreement (Tucker's phi). The Tucker's phi-coefficients for the four race groups are given in Table 7.

TABLE 7
CONSTRUCT EQUIVALENCE OF THE UWES FOR DIFFERENT RACE GROUPS

Group	N	Percentage	Tucker's phi	Tucker's phi	Tucker's phi
White	952	41,55	0,99	0,99	0,99
Black	946	41,29	0,99	0,99	0,99
Coloured	309	13,49	0,99	0,99	0,99
Indian	84	3,67	0,99	0,99	0,99

Inspection of Table 7 shows that the Tucker's phi coefficients for White, Blacks, Coloured and Indian police members were acceptable. Consequently, further bias analyses were carried out on the items of the UWES.

The results of the item bias analyses that were carried out through analysis of variance for the 15 items of the adapted UWES are reported in Table 8.

TABLE 8
ITEM BIAS ANALYSES OF THE UWES

Item	Tot_SS	Df_g	SS_g	F_g	Eta square	Df_i	SS_i	F_i	Eta square
Vigour									
UWES1	5633,00	3	50,50	9,30	0,01	27	71,70	1,50	0,01
UWES8	4868,60	3	280,60	74,20	0,05	27	124,30	3,70	0,03
UWES12	3971,40	3	36,10	9,80	0,01	27	50,00	1,50	0,01
UWES15	4546,00	3	73,70	19,10	0,02	27	101,20	2,90	0,02
UWES17	3698,90	3	20,30	6,00	0,01	27	74,40	2,40	0,02
Dedication									
UWES2	3787,90	3	17,60	6,20	0,01	27	31,40	1,20	0,01
UWES5	3039,20	3	11,50	5,50	0,00	27	31,30	1,70	0,01
UWES7	3720,90	3	8,70	3,50	0,00	27	23,50	1,10	0,01
UWES10	2915,40	3	18,20	8,20	0,01	27	32,30	1,60	0,01
UWES13	3548,00	3	1,20	0,50	0,00	27	17,40	0,70	0,01
Absorption									
UWES3	4320,30	3	113,50	26,50	0,03	27	78,70	2,00	0,02
UWES6	5941,10	3	53,40	9,50	0,01	27	74,60	1,50	0,01
UWES9	3741,70	3	28,90	8,70	0,01	27	40,00	1,30	0,01
UWES11	3975,50	3	28,40	7,90	0,01	27	36,70	1,10	0,01
UWES13	5485,40	3	38,50	7,60	0,01	27	40,20	0,90	0,01

Table 8 shows no practical significant eta square values. This indicates that the means of the race groups for the different score levels do not differ from zero in a systematic way. No uniform or non-uniform bias exist regarding the items of the UWES for Whites, Blacks, Coloureds and Indians. These results provide support for Hypotheses 2.

DISCUSSION

The current study examined, for the first time in South Africa, the psychometric properties of the UWES, an instrument constructed to measure the engagement levels of employees. The objectives were to determine the construct validity and internal consistency of the UWES and to test its construct equivalence and bias for different race groups in a sample of police officers.

In order to obtain a factor structure that best represents the UWES, exploratory factor analysis was used to assess the factorial structure. However, the solution yielded factors that could not be interpreted meaningfully. Because the preliminary research

of Schaufeli and colleagues (2002, in press) concluded that work engagement is a multidimensional construct comprising three dimensions, it was decided to test a three-factor model, using structural equation modelling.

The hypothesised three-factor model of the UWES fitted the data, albeit after removing two unsound items, based on their high standardised residuals, and after allowing some error terms to correlate. The two items that were deleted in the three-factor model were item 4 ("I feel strong and vigorous in my job") and item 14 ("I get carried away by my work").

Because the specification of correlated error terms for purposes of achieving a better-fitting model is not an acceptable practice and error terms were allowed to correlate between items belonging to different subscales (vigour and absorption), the fit of an alternative unidimensional model was assessed as well. This model was also rejected on both substantive and statistical grounds. Additional exploratory work revealed substantial improvement in model fit with the deletion of four items (item 3, "Time flies when I'm working", item 11, "I am immersed in my work", item 15, "I am very resilient, mentally, in my job" and item 16, "It is difficult to detach myself from my job"). Error terms were also allowed to correlate in order to improve model fit (Byrne, 2001).

Although Schaufeli et al. (2002, in press) confirmed a three-dimensional construct in previous studies, the three-factor structure is by no means to be considered self-evident in this sample of police officers. The three-factor model represented the data quite well. However, the one-factor model that included a specification of correlated errors to account for the shared domain-specific variances fitted the data better than the revised three-factor model. This is evident from the lower χ^2 value and goodness-of-fit indexes that indicated better fit, as well as better construct equivalence for the proposed one-factor model.

These results are in contrast to the findings of Schaufeli et al. (in press). Although their hypothesised three-factor model did also not fit well to the data of any of the three samples, the fit of a one-factor model was inferior in comparison with a three-factor model in all three samples. It must be mentioned that they allowed error terms to correlate in all three subscales.

In examining the factor structure, some undesirable psychometric characteristics were found to be associated with several items in the UWES. Items 4 and 14 (in the three-factor model) and items 3, 11, 15, and 16 (in the one-factor model) showed high standardised residual errors. Additionally, these items had the highest modification indexes. These findings suggest that the items may require either deletion or content modification, in which the latter must rather be considered. The particular items may be problematic because they do not correspond to the conceptual domain of the particular dimension (in the case of the three-factor model). However, it is more likely that they are somewhat ambiguous, or that they are either sample- or country-specific. Also, the problems with some of these items may be related to difficult words that some of the participants could have found difficult to understand and/or interpret (e.g. vigorous, immersed and resilient). This is highly likely, because only 11 percent had English as mother tongue.

The prominent correlated errors in this study present an important problem. In general, the specification of correlated error terms for the purpose of achieving a better-fitting model is not an acceptable practice. Correlated error terms in measurement models represent systematic, rather than random, measurement error in item responses. They may derive from characteristics specific either to the items or the respondents (Aish & Jöreskog, 1990). For example, if these parameters reflect item characteristics, they may represent a small omitted factor.

However, as may be the case in this instance, correlated errors may represent respondent characteristics that reflect bias such as yea-/nay-saying, social desirability (Aish & Jöreskog, 1990), as well as a high degree of overlap in item content (when an item, although worded differently, essentially asks the same question) (Byrne, 2001).

However, previous research with psychological constructs in general (e.g. Jöreskog, 1982; Newcomb & Bentler, 1988; Tanaka & Huba, 1984), and with measuring instruments in particular (Byrne, 1988, 2001), has demonstrated that the specification of correlated errors can often lead to substantially better fitting models. Bentler and Chou (1987) also argue that the specification of a model that forces these error parameters to be uncorrelated is rarely appropriate with real data. Therefore, it was considered more realistic to incorporate the correlated errors in this study, rather than to ignore their presence.

It is believed that this confusing state of affairs regarding the UWES does not reflect weaknesses inherent in the instrument, but is rather due to more general factors. First, the UWES is a recently constructed measuring instrument. Therefore, relatively few studies have critically reviewed its psychometric properties. In order to study the construct validity of work engagement in greater detail, additional theory-driven research is needed. Secondly, the UWES is an instrument that was originally constructed from data based on samples of individuals in the Netherlands (Schaufeli & Bakker, 2001). Therefore, valid research that compares levels of work engagement in South Africa is lacking and a thorough psychometric evaluation of this instrument in our specific national context will be influenced by the specific culture of the country (or more specifically, the culture of the police organisation). Schaufeli et al. (in press) also found that the hypothesised three-factor model of work engagement was invariant across Spanish, Dutch and Portuguese samples. Also, the dimensionality of the UWES could be influenced because of the high reported correlations between the three dimensions. Explicit theory indicating exactly how the three sub-scales relate to one another and to other variables must be developed before one can evaluate thoroughly the theoretical validity of a three-component conceptualisation.

Internal consistencies were computed for the three engagement scales, which revealed that all three subscales are sufficiently internally consistent according to the guideline of Nunnally and Bernstein (1994). The alpha coefficient of 0,92 for the one-factor model was considerably higher.

Construct (structural) equivalence was used to compare the factor structures of the UWES for different cultural groups included in the study. Equivalence was acceptable for White, Black, Coloured and Indian police members. Furthermore, bias analyses were carried out on the items of the UWES. Bias was examined for each item separately. In this analysis, it was found that the means of the race groups did not differ in a systematic way. It can be deduced that the UWES items do not show uniform or non-uniform bias. Therefore, it seems acceptable to use the UWES to compare work engagement of different race groups.

In conclusion, the data strongly suggest that the one-factor model better fits the data than the three-factor model. However, there is, as yet, insufficient evidence to suggest that a one-factor model is superior to a three-factor model. Thus, although a one-factor model fits the data better, a three-factor model will also fit the data well. Based on the results obtained in this study, it seems as if the UWES must undergo intensive psychometric evaluation before it could be used as a suitable instrument for measuring engagement of police members in the SAPS.

This study had several limitations. First, self-report measures were exclusively relied upon. This causes a particular problem in

validation studies that use self-report measures exclusively because at least part of the common variance of the measures has to be attributed to method variance (Schaufeli, Maslach & Marek, 1993). The use of a cross-sectional study design also represents a limitation, i.e. that of the ability to test causal assumptions regarding the engagement syndrome. Longitudinal data would allow for forming a better understanding of the true nature of work engagement. Also, items were allowed to correlate in the model specification. This may impose interpretation problems because as correlated error terms are added to the model, the correspondence between the posited construct of interest and the empirically defined factor becomes unclear (Gerbing & Anderson, 1984).

RECOMMENDATIONS

There appear to be several research issues that flow from this study and which require attention in increasing both our understanding of work engagement and the usefulness of this concept. Clearly, further construct validity research is needed to establish more fully the factorial validity of the UWES. None of the solutions could be regarded either as effectively confirming the authors' proposed three-subscale structure, or as an adequate replication of the factor structures found in their studies (Schaufeli et al., 2002, in press).

The second issue relates to problem items. Individual items of the UWES may need to be carefully examined when they are used in South African samples. This issue can also be clarified in future research that compares samples from different occupations. Because different problem items emerged with different models, it is more evident that further construct validity research is needed in order to establish more fully the psychometric soundness of the UWES. The findings of this study also suggest the need for possible improvement to item content. This implies that the wording of certain items must be modified in order to make them more appropriate for the specific context. It also seems important to work towards improving the UWES for South African circumstances by identifying a core set of items that could most validly measure the concept of work engagement.

Five suggestions for future research derive from the present findings. Research is needed to determine the reliability and validity of the UWES in other samples in South Africa. Research is needed in other occupations to establish norms for engagement levels other than police officers. Future studies should use large samples and adequate statistical techniques (e.g. structural equation modelling). Large sample sizes might provide increased confidence that study findings would be consistent across other similar groups. Researchers contemplating future validation of the UWES are urged to utilise statistical programs that can yield a measure of multivariate normality, and provide appropriate estimation procedures, given findings of non-normal data. Fourthly, in order to overcome the problem of systematic measurement error in item responses, it is recommended that the items of the MBI-GS and UWES be combined in a single questionnaire for research purposes. Finally, in future studies structural equation modelling could be used to test the construct equivalence of the UWES. In testing for these equivalencies, sets of parameters (i.e. factor loading paths, factor variances/covariances and structural regression paths) could be tested by increasing restrictions in every step.

REFERENCES

Aish, A.M. & Jöreskog, K.G. (1990). A panel model for political efficacy and responsiveness: An application of LISREL 7 with weighted least squares. *Quality and Quantity*, 19, 716-723.

Arbuckle J.L. (1997). *Amos users' guide version 4.0*. Chicago, IL: Smallwaters Corporation.

Band, S.R. & Manuelle, C.A. (1987). Stress and police officers' performance: An examination of effective coping behavior. *Police Studies*, 10, 122-131.

Bentler, P.M. & Chou, C.P. (1987). Practical issues in structural modelling. *Sociological Methods and Research*, 16, 78-117.

Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136-162). London: Sage.

Byrne, B.M. (1988). The Self Description Questionnaire III: Testing for equivalent factorial validity across ability. *Educational and Psychological Measurement*, 48, 397-406.

Byrne, B.M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications and programming*. Mahwah, NJ: Erlbaum.

Clark, L.A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309-319.

Cleary, T.A. & Hilton, T.L. (1968). An investigation of item bias. *Educational and Psychological Measurement*, 28, 61-75.

Csikszentmihalyi, M. (1990). *Flow. The psychology of optimal experience*. New York: Harper.

Diener, E., Suh, E.M., Lucas, R.E. & Smith, H.I. (1999). Subjective wellbeing: Three decades of progress. *Psychological Bulletin*, 125, 267-302.

Gerbing, D.W. & Anderson, J.C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research*, 11, 572-580.

Greller, M.M., Parsons, C.K. & Mitchell, D.R.D. (1992). Additive effects and beyond: Occupational stressors and social buffers in a police organization. In J.C. Quick, L.R. Murphy & J.J. Hurrell, Jr. (Eds), *Stress and wellbeing at work, assessments and interventions for occupational mental health* (pp. 33-47). Washington, DC: American Psychological Association.

Hart, P.M., Wearing, A.J. & Headey, B. (1995, June). Police stress and wellbeing: Integrating personality, coping and daily work experiences. *Journal of Occupational and Organizational Psychology*, 68, 133-156.

Hoyle, R.H. (1995). The structural equation modeling approach: Basic concepts and fundamental issues. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 1-15). Thousand Oaks, CA: Sage.

Hu, L.T. & Bentler, P.M. (1995). Evaluating model fit. In R.H. Hoyle (Ed), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage.

Hu, L.T. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: A Multidisciplinary Journal*, 6, 1-55.

Jöreskog, K.G. (1982). Analysis of covariance structures. In C. Fornell (Ed.), *A second generation of multivariate analysis Vol 1: Methods* (pp. 200-242). New York: Praeger.

Jöreskog, K.G. & Sörbom, D. (1986). *LISREL user guide version VI* (4th ed.). Mooresville, IL: Scientific Software International.

Jöreskog, K.G. & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications*. Chicago: SPSS, Inc.

Jöreskog, K.G. & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kerlinger, F.N. & Lee, H.B. (2000). *Foundations of behavioral research* (4th ed.). Fort Worth, TX: Harcourt College Publishers.

MacCallum, R.C., Browne, M.W. & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130-149.

Marsh, H.W., Balla, J.R. & Hau, K.T. (1996). An evaluation of Incremental Fit Indices: A clarification of mathematical and empirical properties. In G.A. Marcoulides & R.E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 315-353). Mahwah, NJ: Erlbaum.

Maslach, C. & Leiter, M.P. (1997). *The truth about burnout*. San Francisco: Jossey-Bass.

Maslach, C., Schaufeli, W.B. & Leiter, M.P. (2001). Job burnout. *Annual Review of Psychology*, *52*, 397-422.

Mulaik, S.A., James, L.R., Van Altine, J., Bennett, N., Lind, S. & Stillwell, C.D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105, 430-445.

Myers, D.G. (2000). The funds, friends, and faith of happy people. *American Psychologist*, *55*, 56-67.

Newcomb, M.D. & Bentler, P. M. (1988). Consequences of adolescent drug use: Impact on the lives of young adults. Newbury Park, CA: Sage.

Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Rothmann, S. (2002, March). *Burnout research in South Africa*. Paper presented at the 1st South African Conference on Burnout, Potchefstroom.

SAS Institute. (2000). *The SAS System for Windows: Release 8.01*. Cary, NC: SAS Institute Inc.

Schaufeli, W.B. & Bakker, A.B. (2001). Werk en welbevinden: Naar een positieve benadering in de Arbeids- en Gezondheidspsychologie [Work and wellbeing: Towards a positive occupational health psychology]. *Gedrag en Organizatie*, *14*, 229-253.

Schaufeli, W.B., Maslach, C. & Marek, T. (1993). *Professional burnout: Recent developments in theory and research*. Philadelphia: Taylor and Francis.

Schaufeli, W.B., Salanova, M., González-Romá, V. & Bakker, A.B. (2002). The measurement of engagement and burnout: A confirmative analytic approach. *Journal of Happiness Studies*, *3*, 71-92.

Schaufeli, W.B., Martinez, I., Pinto, A.M., Salanova, M. & Bakker, A.B. (in press). Burnout and engagement in university students: A cross national study. *Journal of Cross Cultural Psychology*.

Schutte, N., Toppinen, S., Kalimo, R. & Schaufeli, W.B. (2000, March). The factorial validity of the Maslach Burnout Inventory-General Survey (MBI-GS) across occupational groups and nations. *Journal of Occupational and Organizational Psychology*, *73*, 53-66.

Seligman, M.E.P. & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, *55*, 5-14.

Shaughnessy, J.J. & Zechmeister, E.B. (1997). *Research methods in psychology* (4th ed.). New York: McGraw-Hill.

Strümpfer, D.J.W. (1995). The origins of health and strength: From "salutogenesis" to "fortigenesis". *South African Journal of Psychology*, *25*, 81-89.

Strümpfer, D.J.W. (2002a). Psychofortology: Review of a new paradigm marching on. *Psychofortoloy* (in press). Available on http://general.rau.ac.za/psych.

Strümpfer, D.J.W. (2002b). *Resilience and burnout: A stitch that could save nine*. Paper presented at the 1st National Burnout Conference, Potchefstroom.

Tanaka, J.S. & Huba, G.J. (1984). Confirmatory hierarchical factor analysis of psychological distress measures. *Journal of Personality and Social Psychology*, *46*, 621-635.

Tucker, L.R. & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrica*, *38*, 1-10.

Van de Vijver, F. & Leung, K. (1997). *Methods and data-analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

Van de Vijver, F. & Tanzer, N.K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, *47*, 263-279.

Watson, D. & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*, 219-235.

West, S.G., Finch, J.F. & Curran, P.J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 65-75). Thousand Oaks, CA: Sage.

Wheaton, B., Muthén, B., Alwin, D.F. & Summers, G.F. (1977). Assessing reliability and stability in panel models. In D.R. Heise (Ed), *Sociological methodology* (pp. 84-136). San Francisco, CA: Jossey-Bass.

Wissing, M.P. & Van Eeden, C. (2002). Empirical clarification of the nature of psychological wellbeing. *South African Journal of Psychology*, *32*, 32-44.

Yung, Y.F. & Bentler, P.M. (1996). Bootstrapping techniques in analysis of mean and covariance structures. In G.A. Marcoulides & R.E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 195-226). Mahwah, NJ: Lawrence Erlbaum Associates.

Zhu, W. (1997). Making bootstrap statistical inferences: A tutorial. *Research Quarterly for Exercise and Sport*, *68*, 44-55.