The Construct Validity of the Utrecht Work Engagement Scale: Multisample and Longitudinal Evidence

Piia Seppälä · Saija Mauno · Taru Feldt · Jari Hakanen · Ulla Kinnunen · Asko Tolvanen · Wilmar Schaufeli

Abstract This study investigated the factor structure and factorial group and time invariance of the 17-item and 9-item versions of the Utrecht Work Engagement Scale (UWES; Schaufeli et al. (2002b) Journal of Happiness Studies 3:71–92). Furthermore, the study explored the rank-order stability of work engagement. The data were drawn from five different studies ($N = 9,404$), including a three-year longitudinal study ($n = 2,555$), utilizing five divergent occupational samples. Confirmatory factor analysis supported the hypothesized correlated three-factor structure—vigor, dedication, absorption—of both UWES scales. However, while the structure of the UWES-17 did not remain the same across the samples and time, the structure of the UWES-9 remained relatively unchanged. Thus, the UWES-9 has good construct validity and use of the 9-item version can be recommended in future research. Moreover, as hypothesized, Structural Equation Modeling showed high rank-order stabilities for the work engagement factors (between 0.82 and 0.86). Accordingly, work engagement seems to be a highly stable indicator of occupational well-being.

Keywords Work engagement · Utrecht Work Engagement Scale · Construct validity · Factor structure · Factorial group and time invariance · Rank-order stability

P. Seppälä (✉) · S. Mauno · T. Feldt · A. Tolvanen
Department of Psychology, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland
e-mail: piia.r.seppala@psyka.jyu.fi

J. Hakanen
Department of Psychology, Finnish Institute of Occupational Health, Helsinki, Finland

U. Kinnunen
Department of Psychology, University of Tampere, Tampere, Finland

W. Schaufeli
Department of Psychology, Utrecht University, Utrecht, The Netherlands

1 Introduction

With the emergence of positive psychology (e.g., Seligman 2002, 2003; Seligman and Csikszentmihalyi 2000; Turner et al. 2002) added to the fact that the number of positive constructs of occupational well-being are limited, the concept of *work engagement* has received increasing attention in the field of occupational health psychology (Schaufeli and Salanova 2007). Work engagement, including the three dimensions of vigor, dedication and absorption, is assumed to be a strictly positive and relatively stable indicator of occupational well-being (Schaufeli et al. 2002b). The three dimensions of work engagement are also included in the *Utrecht Work Engagement Scale* (UWES), a survey which has been developed to measure work engagement (Schaufeli et al. 2002b).

The UWES has been translated into many languages and used among different occupational groups (e.g., blue-collar workers, dentists, hospital staff, managers, police officers, teachers; see Schaufeli 2007a; Schaufeli and Bakker 2003) although its psychometric properties have remained somewhat less explored. For example, it is still unclear whether the theoretically based three-dimensional structure of the scale remains the same across different occupational groups (i.e., factorial group invariance) and/or across different measurement points (i.e., factorial time invariance). Furthermore, while the time-invariance of the structure of the scale is uncertain, the assumed stability of work engagement remains without strong empirical evidence.

Accordingly, the aim of this study was to investigate the psychometric properties of the UWES by utilizing five divergent data sets, one of which was longitudinal, gathered in Finland ($N = 9,404$). Specifically, the purpose of the study was to test the factor structure of the UWES and its group- and time-invariant properties by means of confirmatory factor analysis (CFA). The former was studied by evaluating the factor structure of the UWES across five samples containing different, although mainly white-collar, occupational groups (i.e., dentists, educational staff, health care staff, managers, and young managers), and the latter by utilizing the so-called multi-sample method and among the dentists ($n = 2,555$), three-year longitudinal data with two measurement points. Furthermore, the longitudinal data made it possible to investigate the stability of work engagement during this three-year time-period. Also the stability of work engagement was assessed by using CFA within the Structural Equation Modeling (SEM) framework, a procedure which results in an error-free stability coefficient for work engagement.

1.1 Work Engagement

Work engagement is considered as the positive opposite of burnout (Schaufeli et al. 2002b; see also Maslach et al. 1996; Maslach and Leiter 1997; Maslach et al. 1996, 2001). Specifically, Schaufeli et al. (2002b, p. 74) define work engagement "as a positive, fulfilling, work-related state of mind that is characterized by vigor, dedication, and absorption." *Vigor*, refers to high levels of energy and mental resilience while working, the willingness to invest effort in one's work, and persistence in the face of difficulties. *Dedication* is characterized by a sense of significance, enthusiasm, inspiration, pride, and challenge. *Absorption* refers to being fully concentrated and deeply engrossed in one's work, and is characterized by time passing quickly and difficulties in detaching oneself from work. According to a recent review, work engagement is positively associated, for instance, with mental and psychosomatic health, intrinsic motivation, efficacy beliefs, positive attitudes towards work and the organization, and high performance (Schaufeli and Salanova 2007).

Furthermore, Schaufeli et al. (2002b, p. 74) define work engagement as a relatively stable state of mind: "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 behavior." Thus work engagement is considered to be more stable than work-related emotions (e.g., contented, enthusiastic, cheerful; see Warr 1990), but less stable than personality traits, such as the Big Five (for the distinction between emotions, moods, and temperament; see Gray and Watson 2001). As a matter of fact, work engagement has been considered a work-related mood (Schaufeli and Salanova 2007).

1.1.1 Utrecht Work Engagement Scale

The UWES, a self-report questionnaire, consists of 17 items (UWES-17), which measure the three underlying dimensions of work engagement: vigor (six items), dedication (five items), and absorption (six items) (Schaufeli 2007b; Schaufeli et al. 2002b; see Appendix). At first the UWES consisted of 24 items, but after psychometric testing seven unsound items were omitted and 17 items were retained. Subsequent psychometric analyses revealed another two weak items (item 6 in the scale of vigor and item 6 in the scale of absorption; see Schaufeli and Bakker 2003), and hence a 15-item version of the UWES has been used in some studies (e.g., Xanthopoulou et al., in press). Recently, a shorter 9-item version of the UWES (UWES-9) has also been developed (Schaufeli et al. 2006; see Appendix). In this abridged scale, vigor, dedication and absorption are assessed by three items per dimension.

1.1.2 Previous Studies of the UWES

Recent confirmatory factor analytic (CFA) studies have supported the theoretically based correlated three-factor—vigor, dedication, absorption—structure of the UWES-17 and UWES-9 (e.g., Hakanen 2002; Hallberg and Schaufeli 2006; Schaufeli and Bakker 2003; Schaufeli et al. 2002b, 2006). All these studies have also shown that the three factors of work engagement are highly interrelated (correlations between 0.60 and 0.99). Because of the high correlations between the three factors, an alternative one-factor structure of the UWES-17 and the UWES-9 has also been tested. In this one-factor structure all the items were constrained to load on one underlying factor (Hallberg and Schaufeli 2006; Schaufeli and Bakker 2003; Schaufeli et al. 2002b, 2006). However, in all these studies, the theoretically based correlated three-factor structure has shown significantly better fit with the data than an alternative one-factor structure and thus has received most support.

In addition to verifying the theoretically based structure of the scale, it is very important to determine whether the structure of the scale remains the same (i.e., at least the factor loadings remain equal; see, e.g., Jöreskog 2005; see also Statistical analyses below) across different contexts and over time. If not, we cannot be sure that we are measuring the same construct and what the construct we are measuring actually is. Therefore, to obtain comparable results the construct (the structure of the scale) needs to be the same despite, for example, occupation, culture, or time point.

Thus far, studies of the factorial invariance of the correlated three-factor structure of the UWES-17 and the UWES-9 across groups (i.e., factorial group invariance) have remained rather limited. In fact, the factorial group invariance of the UWES-17 has not been fully confirmed, as in previous group-invariance studies the scale has been revised. In a study conducted among students from Spain, the Netherlands, and Portugal, the correlated three-factor structure of a 14-item UWES-S (i.e., a slightly shortened student version of the

UWES-17; vigor item 6 and absorption items 4 and 6 were removed because of non-significant or poor (<0.40) factor loadings) was found to be only partially group-invariant (see Schaufeli et al. 2002a). The unconstrained correlated three-factor structure showed significantly better fit with data than the constrained correlated three-factor structures (i.e., factor loadings and error covariances were constrained to be equal) in all pairs of countries. The detailed analysis also showed that from one to three of the factor loadings differed between the countries; only the factor loadings of the absorption subscale remained the same across all the country comparisons, while the factor loadings of the vigor subscale remained the same in two of the three countries.

In addition, a study conducted among Greek and Dutch employees failed to support the factorial group invariance of the correlated three-factor structure of the UWES-15 (i.e., vigor item 6 and absorption item 6 were removed). In particular, the unconstrained correlated three-factor structure showed significantly better fit with data than the constrained correlated three-factor structures (i.e., factor loadings, factor variances, and error variances were constrained to be equal) (Xanthopoulou et al., in press). To date, only Schaufeli et al. (2006) have tested the group-invariance of the UWES-9. They found that the correlated three-factor structure of the UWES-9 did not remain the same across 10 countries (Australia, Belgium, Canada, Finland, France, Germany, The Netherlands, Norway, South Africa, Spain). Specifically, the unconstrained correlated three-factor structure fitted the data significantly better than the constrained correlated three-factor structures (i.e., the factor loadings and factor covariances were constrained to be equal across 10 countries). Taken together, the results on factorial group-invariance are somewhat conflicting and the equality of the correlated three-factor structure of the UWES-17 and the UWES-9 in different contexts has not been shown.

Longitudinal studies of work engagement are rare, and to the best of our knowledge the factorial invariance of the correlated three-factor structure of the UWES-17 and the UWES-9 across time (i.e., factorial time invariance) has not previously been studied (see Schaufeli 2007a). Therefore, it continues to be unclear whether the hypothesized correlated three-factor structure of the UWES-17 and the UWES-9 remains unchanged over time. Moreover, the hypothesized stability of work engagement remains unclear, since the stability of the construct can only be evaluated if the structure of the scale remains unchanged.

In the present study, CFA within the SEM framework makes it possible to test the invariance of the correlated three-factor structure of the UWES-17 and the UWES-9 across time by connecting CFA models tested at different measurement points in the same model (e.g., Jöreskog 2005). Furthermore, demonstration of the structural time-invariance of the UWES-17 and the UWES-9 allows for the production of error-free rank-order stability coefficients of work engagement. Rank-order stability reflects the degree to which the relative ordering of individuals within a group is maintained over time. Therefore, rank-order stability is conceptually and statistically distinct from traditional exploratory techniques, such as investigating absolute mean level changes occurring in a concept over time.

Thus far, the rank-order stability of work engagement has been assessed by using correlation coefficients between the corresponding constructs (Llorens et al. 2007; Mauno et al. 2007; Schaufeli et al. 2006). In previous longitudinal studies, the experience of work engagement has tended to remain fairly stable. In a two-year follow-up study in Finland, the test-retest correlations of the UWES-17 for vigor, dedication and absorption were 0.73, 0.67, 0.69, respectively (Mauno et al. 2007). Also, in a one-year follow-up study conducted in Australia and Norway, the corresponding test-retest correlations of the UWES-9 were 0.61, 0.56, and 0.60 for Australia, and 0.71, 0.66, and 0.68 for Norway (Schaufeli et al. 2006). In a study among Spanish university students, the test-retest correlations were

0.68 for vigor and 0.61 for dedication over a three-week period, when a slightly adapted student version of the vigor and dedication subscales of the UWES-17 were used in a laboratory setting (Llorens et al. 2007).

1.2 Aims of the Study

To sum up, the psychometric testing of the UWES is still in progress and warrants further research. So far, only a few studies have been conducted on factorial group invariance (Schaufeli et al. 2002a, 2006; Xanthopoulou et al., in press). These have shown somewhat conflicting results and in no case has factorial group invariance been demonstrated. Moreover, according to the definition, work engagement is considered a stable rather than a momentary state of mind (Schaufeli et al. 2002b); however, the true stability of work engagement remains unclear, as evidence for the factorial time invariance of the scale is lacking. In addition, the short version of the scale has only recently been developed, and therefore the psychometric properties of the UWES-9 have not yet been fully tested. The present study addresses each of these concerns.

Specifically, the present study investigated the construct validity of the Finnish translations of the UWES-17 and the UWES-9. The first aim was to test whether the Finnish translations of the scales would include the three interrelated theoretically based dimensions of vigor, dedication, and absorption. To ensure the validity of the hypothesized structure, and since the three dimensions of work engagement have correlated highly in previous studies, an alternative one-factor structure of the UWES-17 and UWES-9 was also tested. The second aim was to investigate whether the correlated three-factor structure would remain the same across different occupational samples (i.e., factorial group invariance). The third aim of this study was to investigate whether the correlated three-factor structure would remain the same across different measurement points (i.e., factorial time invariance) by conducting a three-year follow-up study among dentists. The final aim was to examine the rank-order stabilities of the three work engagement factors across the three-year follow-up period.

The study hypotheses, on the basis of the theory and previous studies of work engagement, can be summarized as follows:

H1: The correlated three-factor structure of the UWES-17 and the UWES-9 fits better to the data than the one-factor structure in each occupational sample and at both measurement points.
H2: The correlated three-factor structure of the UWES-17 and the UWES-9 remains the same across the five occupational samples.
H3: The correlated three-factor structure of the UWES-17 and the UWES-9 remains the same across the two measurement points.
H4: The rank-order stabilities of the work engagement factors are relatively high over the three-year follow-up time.

2 Method

2.1 Participants

The study was based on five independent samples containing a total of 9,404 Finnish participants. Table 1 shows the distribution of background factors (gender, age,

Table 1 Background characteristics and mean levels of the dimensions of work engagement for the study samples

Variable	Sample 1 Health care (n = 736)	Sample 2 Young managers (n = 747)	Sample 3 Managers (n = 1,301)	Sample 4 Education (n = 3,365)	Sample 5 Dentists (n = 3,255, follow-up n = 2,555) 2003	2006	Total N = 9,404
Gender							
Female	639 (87%)	108 (15%)	394 (30%)	2,534 (79%)	2,328 (72%)	1,883 (74%)	6,003 (65%)
Male	97 (13%)	637 (85%)	902 (70%)	676 (21%)	927 (28%)	672 (26%)	3,239 (35%)
Age	M = 44.00 SD = 9.81 Range = 19–63	M = 30.96 SD = 3.22 Range = 23–35	M = 47.88 SD = 8.52 Range = 23–65	16–25 (5%) 26–35 (24%) 36–45 (25%) 46–55 (31%) Over 55 (15%)	M = 45.79 SD = 9.43 Range = 23–82	M = 48.49 SD = 8.82 Range = 27–76	Range = 16–82
Organizational/job tenure	Organizational tenure	Organizational tenure	Organizational tenure	Job tenure	Job tenure		Range = 0–52
	M = 13.81 SD = 9.38 Range = 0–40	M = 4.37 SD = 3.62 Range = 0–18	M = 11.86 SD = 9.35 Range = 0–40	M = 12.37 SD = 9.86 Range = 0–41	M = 19.54 SD = 9.95 Range = 0–50	M = 22.30 SD = 9.41 Range = 3–52	
Permanent job	579 (80%)	696 (93%)	1,261 (98%)	2,183 (67%)	2,998 (94%)	2,404 (96%)	7,717 (84%)
Fixed-term job	148 (20%)	50 (7%)	29 (2%)	1,087 (33%)	200 (6%)	87 (4%)	1,514 (16%)
Hours worked weekly	M = 38.43 SD = 6.07 Range = 8–85	M = 38.52 SD = 2.90 Range = 7–50	M = 47.28 SD = 8.33 Range = 16–90	M = 36.77 SD = 9.52 Range = 2–75	M = 32.88 SD = 8.16 Range = 0–79	M = 32.45 SD = 7.81 Range = 2–55	Range = 0–90
UWES-17							
Vigor	M = 4.51, SD = 1.05	M = 4.57, SD = 0.93	M = 4.52, SD = 0.97	M = 4.58, SD = 1.01	M = 4.52, SD = 1.04	M = 4.59, SD = 0.95	Range = 4.51–4.59
Dedication	M = 4.82, SD = 1.12	M = 4.63, SD = 1.13	M = 4.80, SD = 1.08	M = 4.67, SD = 1.19	M = 4.97, SD = 1.09	M = 4.97, SD = 0.99	Range = 4.63–4.97
Absorption	M = 3.82, SD = 1.38	M = 4.07, SD = 1.21	M = 4.34, SD = 1.15	M = 3.92, SD = 1.34	M = 3.81, SD = 1.40	M = 3.81, SD = 1.35	Range = 3.81–4.34
UWES-9							

Table 1 continued

Variable	Sample 1 Health care (n = 736)	Sample 2 Young managers (n = 747)	Sample 3 Managers (n = 1,301)	Sample 4 Education (n = 3,365)	Sample 5 Dentists (n = 3,255, follow-up n = 2,555)		Total N = 9,404
Vigor	M = 4.64, SD = 1.17	M = 4.55, SD = 1.08	M = 4.53, SD = 1.11	M = 4.56, SD = 1.18	M = 4.63, SD = 1.24	M = 4.71, SD = 1.14	Range = 4.53–4.71
Dedication	M = 4.66, SD = 1.27	M = 4.54, SD = 1.19	M = 4.71, SD = 1.15	M = 4.56, SD = 1.29	M = 4.89, SD = 1.22	M = 4.85, SD = 1.15	Range = 4.54–4.89
Absorption	M = 4.10, SD = 1.40	M = 4.13, SD = 1.29	M = 4.41, SD = 1.18	M = 4.13, SD = 1.44	M = 4.22, SD = 1.44	M = 4.23, SD = 1.37	Range = 4.10–4.41

organizational/job tenure, type of job contract, hours worked weekly) among the partici-
pants and the mean levels for the dimensions of work engagement in each sample.

Sample 1 consisted of questionnaire data collected in a single public health care
organization in Finland in 2003 (Central Finland Health Care District, henceforth health
care; see Mauno et al. 2005, 2007). Of the selected random sample, 736 participants
returned the questionnaire (response rate 46%). The majority of the participants were
women (87%) and worked as nurses ($n = 468$; 64%). The other occupational groups
represented in the sample were physicians ($n = 38$; 5%), clerical workers ($n = 85$; 12%)
(i.e., support services), administrative ($n = 77$; 11%), researchers/research assistants
($n = 37$; 5%), and technical/warehouse/delivery workers ($n = 23$; 3%).

Sample 2 consisted of questionnaire data gathered in 2006 from registered members of
two Finnish trade unions (Union of Professional Engineers in Finland and Union of Sal-
aried Employees). A postal questionnaire was sent to all members of the trade unions who
were age 35 years or less and in a managerial position (henceforth young managers). The
sample consisted of 747 managers (response rate 49%). The great majority of the partic-
ipants were men (86%) and they worked in different parts of Finland in both the private
and public sectors. Of them 43% represented lower management, 49% middle management
and 8% top management.

Sample 3 consisted of questionnaire data gathered in 2005 from the members of five
large Finnish trade unions (Union of Professional Engineers in Finland, Finnish Associ-
ation of Graduates in Economics and Business, Finnish Association of Graduate Engineers,
Finnish Association for Human Resource Management, and Experts and Managerial
Professionals of Municipalities Association; see Kinnunen et al. 2008). The random
sample consisted of 1,301 participants (response rate 40%) after omitting those employees
who were not in a managerial position (henceforth managers). The majority of the par-
ticipants were men (70%), and 42% represented lower management, 26% middle
management and 32% top management.

Sample 4 consisted of questionnaire data gathered in 2001 from the Educational
Department of Helsinki, Finland (henceforth education; see Bakker et al. 2007; Hakanen
2002; Hakanen et al. 2006). The sample consisted of 3,365 participants (response rate
52%) from across the whole organization and from all the professional groups within it.
The majority of participants were female (79%) and most of them worked as teachers
($n = 2,038$; 60%) at elementary ($n = 843$), lower secondary ($n = 497$), upper secondary
($n = 278$), or vocational schools ($n = 217$). The other occupational groups represented in
the sample were support staff (e.g., psychologist, school assistant) ($n = 936$; 28%) and
administrative workers ($n = 391$; 12%).

Sample 5 consisted of questionnaire data collected in a three-year follow-up study
(2003–2006) with two measurement points among Finnish dentists (henceforth dentists;
see Hakanen et al. 2005). The postal questionnaire was sent to every dentist who was a
member of the Finnish Dental Association at the time the data were gathered in 2003. In
2003, 3,255 dentists answered the questionnaire (response rate 71%), and in 2006, 2,555 of
those identified three years later ($n = 3,035$) returned the questionnaire (response rate
84%). Most of the respondents (over 70%) were women and 60% of the respondents were
employed in the public sector.

2.2 Instrument

Work Engagement was assessed by using Finnish translations of the UWESs 17 and 9
(Hakanen 2002). The accuracy of the Finnish translations was checked by the back-

translation method. The UWES-17 consists of 17 items on the three underlying dimensions of vigor, dedication, and absorption (Schaufeli 2007b; see Appendix). "Vigor" is measured with six items (items 1, 4, 8, 12, 15, 17), "dedication" with five items (items 2, 5, 7, 10, 13), and "absorption" with six items (items 3, 6, 9, 11, 14, 16). Items are rated on a seven-point scale ranging from 0 (*never*) to 6 (*every day*). The internal consistencies (Cronbach's alpha) of the UWES-17 ranged between 0.75 and 0.83 for vigor, between 0.86 and 0.90 for dedication, and between 0.82 and 0.88 for absorption. The shortened version of the UWES, the UWES-9, contains three items for vigor (items 1, 4, 8), three for dedication (items 5, 7, 10) and three for absorption (items 9, 11, 14) (Schaufeli et al. 2006; see Appendix). The items are rated on the same scale as the UWES-17. The internal consistencies (Cronbach's alpha) of the UWES-9 varied from 0.81 to 0.85 for vigor, from 0.83 to 0.87 for dedication, and from 0.75 to 0.83 for absorption.

2.3 Statistical Analyses

To investigate the psychometric properties of the UWES scales, CFA within the SEM framework, performed with the LISREL 8.72 program (Jöreskog and Sörbom 1996a), was used. As the variables were ordinal, the Weighted Least Squares (WLS) estimation procedure based on polychoric correlations and asymptotic covariance matrices, calculated by the PRELIS 2.72 program (Jöreskog and Sörbom 1996b), was applied. Likewise, as the variables were ordinal, threshold values of the observed variables were set to be equal for each sample and in Sample 5 for both measurement points, using the PRELIS 2.72 program (Jöreskog and Sörbom 1996b). This was done to ensure that the response scale was the same across the investigated samples and across the different measurement points (see, e.g., Jöreskog 2005). The distributions of the responses were skewed—the majority of the answers were in categories 4 ("*Once a week*") or 5 ("*A few times a week*"). The polychoric correlation is, however, considered to be rather robust to violations of underlying bivariate normality (see, e.g., Jöreskog 2005). Because there is no consensus on the fit indices for evaluating structural equation models (e.g., Bollen and Long 1993; Boomsma 2000; Hoyle and Panter 1995), model fit and model comparisons were based on several fit indices (see Fit indices below). This study focused on the participants who answered all the items of the UWES-17 and the UWES-9 (i.e., listwise deletion).

The analytic procedure consisted of three steps. In the first step of the analyses, the hypothesized correlated three-factor model (henceforth M1; see Fig. 1a) of the UWES-17 and the UWES-9 was tested for each sample separately, and also for both measurement points in Sample 5, to determine whether the observed variables (items) of work engagement loaded on the hypothesized latent factors (vigor, dedication, absorption) on each occasion. To ensure that the correlated three-factor model was valid, an alternative one-factor model (henceforth M2; see Fig. 1b) of the UWES-17 and the UWES-9 was also tested. In the one-factor model, all the items were restricted to load on one latent factor.

In the second step of the analyses, the factorial group invariance of the best-fitting factor model was simultaneously investigated across the five samples by using the multi-sample method (i.e., data on the same variables collected from several samples were estimated in a joint analysis). The factorial group invariance of the UWES-17 and the UWES-9 was tested by comparing the fit of the baseline model (i.e., thresholds of observed variables were constrained to be equal and other parameter estimates were freely estimated) to that of the constrained model (i.e., thresholds of observed variables and factor loadings were constrained to be equal across the five samples). Since the statistical power of a structural
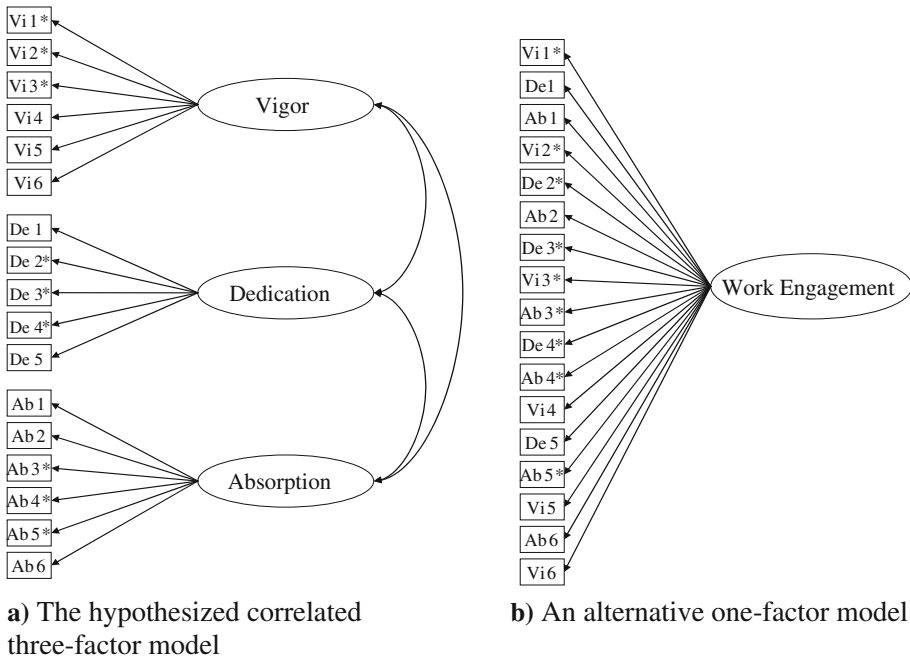
a) The hypothesized correlated three-factor model

b) An alternative one-factor model

Fig. 1 The CFA models of the UWES-17 and the UWES-9 tested in the study (* = UWES-9)

equation model is a function of the characteristics of the model and size of the sample, the large sample size of this study might produce results that are of statistical, but not practical, significance (e.g., Hoyle and Panter 1995; see also Fit indices below). Therefore, because of the very large sample size ($N = 9,404$), only the weak measurement invariance was tested. To determine weak measurement invariance, the equality of the factor loadings (i.e., same unit of measurement; see Jöreskog 2005) across samples must be demonstrated; this is not required of the other parameters (see Meredith 1964, 1993). In other words, the equality of the factor loadings is the minimum assumption for factorial group invariance to hold, because if the factor loadings do not remain the same across different groups, we cannot be sure that we are measuring the same construct and what the construct we are measuring actually is.

In the third step of the analyses, the factorial time invariance of the best-fitting factor model of the UWES-17 and the UWES-9 was investigated over a three-year follow-up with two measurement points (Sample 5). First, the baseline stability model was estimated by using structural equations between the latent factors of work engagement to connect the factor models estimated at the two time points. Next, the constrained stability model was estimated by setting the corresponding factor loadings equal across the two measurement points. Factorial time invariance was then tested by comparing the fit of the baseline stability model (i.e., thresholds of observed variables were set equal and other parameters freely estimated) to the constrained stability model (i.e., thresholds of observed variables and factor loadings were set equal across the two measurement points). As previously, the equality of the factor loadings at different measurement points must be demonstrated; however, equality of the other parameters is not required to determine weak measurement invariance (see Meredith 1964, 1993). Finally, the rank-order stabilities of the work engagement factors were investigated by estimating the stability coefficients of the UWES-17

and the UWES-9 in Sample 5 (i.e., β-coefficients between the three latent factors of work engagement at Time 1 and Time 2).

2.3.1 Fit Indices

The fit of the different models was evaluated by using several types of fit indices. The absolute fit of the model was assessed with the chi-square index and the fit of the competing models was compared with the chi-square difference test (Satorra–Bentler's scaling corrections were used; Satorra and Bentler 2001). A general rule is that a non-significant chi-square value indicates good model fit. In addition, in the chi-square difference test a non-significant reduction in the chi-square, relative to the change in the number of degrees of freedom, indicates that the constrained model is acceptable. If the reduction in the chi-square is significant, the baseline model is more satisfactory. However, a well-known disadvantage of the chi-square statistic is its high sensitivity to sample size (e.g., Bentler and Bonett 1980). With large sample sizes most models tend to be rejected, although deviations of the empirical data from theoretical expectations (i.e., chi-square test) or a significant reduction in the chi-square value for the constrained model (i.e., chi-square difference test) are for practical purposes irrelevant and due to the overwhelming statistical power of the sample size (e.g., Bentler and Bonett 1980). Therefore, the fit and the invariance of the models were also evaluated by other fit indices. The RMSEA (Root Mean Square Error of Approximation) is also an absolute fit index and it indicates the error of approximation; values of 0.05 or less indicate close fit of the model, values below 0.08 indicate reasonable model fit, and values over 0.1 indicate poor model fit (e.g., Browne and Cudeck 1993; Yu 2002). The incremental fit of the models was evaluated by using the CFI (Comparative Fit Index) and the NNFI (Non Normed Fit Index). The CFI and NNFI measure the improvement in the fit by comparing the hypothesized model with an independence model that specifies no covariances among the variables; the CFI and NNFI values should be 0.95, or preferably, above, to indicate good model fit (e.g., Hu and Bentler 1999; Yu 2002).

3 Results

3.1 The Factor Structure of the UWES-17 and the UWES-9

The goodness-of-fit statistics for the hypothesized correlated three-factor model (M1) and an alternative one-factor model (M2) of the UWES-17 and the UWES-9 are summarized in Tables 2 and 3. As Tables 2 and 3 show, M1 fitted—except according to the chi-square test—the data well, but the fit of M2 was also acceptable. M1, however, showed a better fit with the data than M2. The chi-square difference test also revealed that the fit of M1 was significantly better than that of M2 in all five samples and at both measurement points (H1). Thus, work engagement as a three-dimensional construct gained support and M1 was chosen for the subsequent analyses of factorial group and time invariance. As expected, the three latent factors were highly correlated. The correlations between the latent factors of the UWES-17 ranged from 0.90 to 0.96, and the correlations between the latent factors of the UWES-9 ranged from 0.83 to 0.97.

However, it should be noted that although the overall fit of M1 of the UWES-17 was good, absorption item 6 ("*It is difficult to detach myself from my job*") showed a somewhat

Table 2 Goodness-of-fit statistics for the alternative CFA models of the UWES-17

Sample	Models	$\chi^2(df)$	RMSEA	CFI	NNFI	$\Delta\chi^2(\Delta df)$
Sample 1						
($n = 674$)	M1	409.06 (116)	0.061	0.97	0.96	
($n = 674$)	M2	471.41 (119)	0.066	0.96	0.95	62.35 (3), $p < 0.001$
Sample2						
($n = 730$)	M1	479.59 (116)	0.066	0.97	0.97	
($n = 730$)	M2	567.15 (119)	0.072	0.97	0.96	87.56 (3), $p < 0.001$
Sample 3						
($n = 1,275$)	M1	771.23 (116)	0.067	0.97	0.96	
($n = 1,275$)	M2	858.89 (119)	0.070	0.96	0.96	87.66 (3), $p < 0.001$
Sample 4						
($n = 2,971$)	M1	1392.14 (116)	0.061	0.94	0.93	
($n = 2,971$)	M2	1616.29 (119)	0.065	0.93	0.92	224.15 (3), $p < 0.001$
Sample 5						
($n = 2,723$)	M1	1524.97 (116)	0.067	0.94	0.93	
($n = 2,723$)	M2	1685.64 (119)	0.070	0.94	0.93	160.67 (3), $p < 0.001$
Follow-up ($n = 2,314$)	M1	1311.81 (116)	0.067	0.94	0.92	
Follow-up ($n = 2,314$)	M2	1483.74 (119)	0.070	0.93	0.92	171.93 (3), $p < 0.001$

M1 = hypothesized correlated three-factor model

M2 = alternative one-factor model

lower factor loading in one sample and vigor item 6 ("*At my work I always persevere, even when things do not go well*") showed slightly lower factor loadings in two samples when compared to the other factor loadings. For absorption item 6, the factor loading was 0.51 in Sample 5, and for vigor item 6, the factor loading was 0.37 in Sample 5 and 0.56 in Sample 4. The other factor loadings were in general high, ranging from 0.61 to 0.99.

With respect to the UWES-9 it should be pointed out that the RMSEA values of M1 did not meet the cut-off criteria in all samples. On the basis of the information given by the modification indices, absorption item 4 ("*I am immersed in my work*") and absorption item 5 ("*I get carried away when I'm working*") revealed a rather strong correlation between their error variances in all five samples (error covariances varied between 0.18 and 0.29). As the UWES-9 has only recently been developed and thus has been less studied, a modified correlated three-factor model ($M1_{mod.}$), in which the error variances of absorption items 4 and 5 were allowed to correlate, was also computed. This led to an improvement in the fit of the model (see $M1_{mod.}$ in Table 3). However, as the viewpoint of the study was confirmatory and as the hypothesized model needed only this minor modification, the comparisons between the alternatively restricted models were done with the original M1.

3.2 The Factorial Group Invariance of the UWES-17 and the UWES-9

Table 4 reports the results for the factorial group invariance tests of M1 of the UWES-17 and the UWES-9. As Table 4 shows, the baseline model of the UWES-17 showed good fit—except according to the chi-square test—with the data. However, on the basis of the chi-square difference test, the factor loadings invariance assumption was not supported, as the constrained model displayed a significant loss of fit when compared to the baseline

Table 3 Goodness-of-fit statistics for the alternative CFA models of the UWES-9

Sample	Models	χ^2(df)	RMSEA	CFI	NNFI	$\Delta\chi^2(\Delta df)$
Sample 1						
($n = 699$)	M1	124.52 (24)	0.077	0.98	0.96	
($n = 699$)	M1$_{mod.}$	74.48 (23)	0.057	0.99	0.98	
($n = 699$)	M2	150.25 (27)	0.081	0.97	0.96	25.73 (3), $p < 0.001$
Sample 2						
($n = 732$)	M1	144.19 (24)	0.083	0.98	0.97	
($n = 732$)	M1$_{mod.}$	117.63 (23)	0.075	0.98	0.97	
($n = 732$)	M2	176.92 (27)	0.087	0.97	0.96	32.73 (3), $p < 0.001$
Sample 3						
($n = 1,288$)	M1	268.87 (24)	0.089	0.97	0.96	
($n = 1,288$)	M1$_{mod.}$	217.38 (23)	0.081	0.98	0.97	
($n = 1,288$)	M2	355.68 (27)	0.097	0.96	0.95	86.81 (3), $p < 0.001$
Sample 4						
($n = 3,140$)	M1	348.09 (24)	0.066	0.98	0.97	
($n = 3,140$)	M1$_{mod.}$	204.23 (23)	0.050	0.99	0.98	
($n = 3,140$)	M2	530.48 (27)	0.077	0.96	0.95	182.39 (3), $p < 0.001$
Sample 5						
($n = 2,935$)	M1	458.18 (24)	0.079	0.97	0.96	
($n = 2,935$)	M1$_{mod.}$	217.95 (23)	0.054	0.99	0.98	
($n = 2,935$)	M2	669.68 (27)	0.090	0.96	0.95	211.50 (3), $p < 0.001$
Follow-up ($n = 2,399$)	M1	391.85 (24)	0.080	0.97	0.95	
Follow-up ($n = 2,399$)	M1$_{mod.}$	146.11 (23)	0.047	0.99	0.98	
Follow-up ($n = 2,399$)	M2	574.99 (27)	0.092	0.95	0.93	183.14 (3), $p < 0.001$

M1 = hypothesized correlated three-factor model

M1$_{mod.}$ = modified correlated three-factor model, error variances of absorption items 4 and 5 were allowed to correlate

M2 = alternative one-factor model

Table 4 Group-invariance tests for the correlated three-factor model of the UWES-17 and the UWES-9

Multi-sample comparisons	χ^2	df	$\Delta\chi^2(\Delta df)$	p	RMSEA	CFI	NNFI
1. Baseline model							
UWES-17	4541.55	580			0.064	0.96	0.95
UWES-9	1328.65	120			0.076	0.98	0.96
2. Constrained model							
UWES-17	5293.46	636	2 vs. 1		0.066	0.95	0.94
			751.91 (56)	< 0.001			
UWES-9	1432.69	144	2 vs. 1		0.071	0.97	0.97
			104.04 (24)	< 0.001			

Baseline model = thresholds of observed variables set to be equal across five samples

Constrained model = thresholds of observed variables and factor loadings set to be equal across five samples

model. This indicates that the size of the factor loadings was not equal but differed across the samples, which led to the deterioration in the model fit when the factor loadings were constrained to be equal. Although the chi-square difference test is highly sensitive to sample size, the loss of fit regarding degrees of freedom was notable (χ^2/df ratio = 13.4). The χ^2/df ratio estimates how many times larger the chi-square estimate is than its expected value, and ratios indicating good fit range from 2 to 5 (see Bollen 1989). Thus, it is possible that even with a smaller sample size the chi-square difference test would not have supported the group invariance assumption.

The detailed analysis of the paired comparisons of the samples also revealed that the factor loadings invariance assumption was not supported in any of the pairs of samples (see Table 5). The constrained model displayed a significant loss of fit when compared to the baseline model in all pairs of samples, indicating that the size of the factor loadings was different in every pair of the samples. In addition, the detailed analysis of the factor loadings of the UWES-17 revealed marked variation in the factor loadings across the five samples, varying from 0.03 to 0.44 (on the basis of M1 with all parameters estimated freely, completely standardized solution). The largest loading variations were for the

Table 5 Paired comparisons of the samples for the correlated three-factor model of the UWES-17 and the UWES-9

Combination of samples	Baseline model $\chi^2(df)$	Constrained model $\chi^2(df)$	$\Delta\chi^2(\Delta df)$	p
UWES-17				
1, 2	872.18 (232)	914.06 (246)	41.88 (14)	< 0.001
1, 3	1151.10 (232)	1195.59 (246)	44.49 (14)	< 0.001
1, 4	1790.96 (232)	1866.52 (246)	75.56 (14)	< 0.001
1, 5	1924.65 (232)	2111.44 (246)	186.79 (14)	< 0.001
2, 3	1242.91 (232)	1287.90 (246)	44.99 (14)	< 0.001
2, 4	1878.67 (232)	2067.98 (246)	189.31 (14)	< 0.001
2, 5	2002.24 (232)	2392.88 (246)	390.64 (14)	< 0.001
3, 4	2148.09 (232)	2315.54 (246)	167.45 (14)	< 0.001
3, 5	2272.37 (232)	2695.22 (246)	422.85 (14)	< 0.001
4, 5	2922.13 (232)	3078.94 (246)	156.81 (14)	< 0.001
UWES-9				
1, 2	260.08 (48)	276.93 (54)	16.85 (6)	= 0.010
1, 3	388.03 (48)	397.28 (54)	9.25 (6)	= 0.160
1, 4	472.76 (48)	487.91 (54)	15.15 (6)	= 0.019
1, 5	584.06 (48)	592.13 (54)	8.07 (6)	= 0.233
2, 3	404.87 (48)	411.70 (54)	6.83 (6)	= 0.337
2, 4	489.70 (48)	528.01 (54)	38.31 (6)	< 0.001
2, 5	600.32 (48)	628.15 (54)	27.83 (6)	< 0.001
3, 4	606.91 (48)	635.58 (54)	28.67 (6)	< 0.001
3, 5	712.08 (48)	734.74 (54)	22.66 (6)	< 0.001
4, 5	807.19 (48)	861.25 (54)	54.06 (6)	< 0.001

Baseline model = thresholds of observed variables set to be equal across pair of samples

Constrained model = thresholds of observed variables and factor loadings set to be equal across pair of samples

loadings of absorption item 6 and vigor item 6. Finally, on the basis of the modification indices seven factor loadings were estimated as unequal (loadings of vigor items 4, 5, 6; dedication item 2; absorption items 2, 3, 6). However, even after these loading modifications, the chi-square difference test still produced a significant loss of fit for the constrained model when compared to the baseline model [$\Delta\chi^2$ (28) = 170.13]. Thus, the UWES-17 measured work engagement differently, according to the occupational group, and the group-invariance hypothesis (H2) regarding the UWES-17 was not supported.

Table 4 also shows the results for the corresponding factorial group invariance tests of the UWES-9. The baseline model of the UWES-9 also showed good fit—except according to the chi-square test—with the data. However, on the basis of the chi-square difference test the factor loadings invariance assumption was not supported as the constrained model displayed a significant loss of fit when compared to the baseline model. As before, this indicates that the size of the factor loadings differed across the samples. However, the loss of fit regarding degrees of freedom was rather modest (χ^2/df ratio = 4.3; see Bollen 1989), and therefore it is possible that the significant reduction in the chi-square value is related to the large sample size.

The detailed analysis of the paired comparisons of the samples also revealed that the factor loadings had different degrees of invariance, depending on the sample size (see Table 5). The factor loadings invariance assumption was mainly supported among the pairs of small samples (i.e., $n < 1,000$) and on those occasions where pairs of small and large (i.e., $n \geq 1,000$) samples were combined, except for pairs 2 and 4, and 2 and 5. On the other hand, for the constrained model the loss of fit was significant among the pairs of large samples. In addition, the detailed analysis of the factor loadings of the UWES-9 showed only modest variation in the factor loadings across the five samples, varying from 0.02 to 0.11 (on the basis of M1 with all parameters estimated freely, completely standardized solution). Except for absorption item 4, the variation in the factor loadings was less than 0.10. Therefore, the factor loadings of absorption item 4 were estimated as unequal, but this modification hardly at all affected the value of the chi-square difference test [$\Delta\chi^2$ (20) = 88.86].

To sum up: on the basis of the chi-square difference test, the factor loadings invariance assumption of the UWES-9 was rejected; however, the detailed analyses revealed that the assumption of the invariance of the factor loadings was supported among the small samples and also showed only minor variation in the factor loadings across the samples. Therefore, the UWES-9 measured work engagement rather similarly among the different occupations and the group-invariance hypothesis (H2) regarding the UWES-9 was relatively well supported.

3.3 The Factorial Time Invariance and Rank-order Stability of the UWES-17 and the UWES-9

Table 6 reports the results for the factorial time invariance tests of M1 of the UWES-17 and the UWES-9. The fit of the baseline stability model of the UWES-17 was relatively good, although the modification indices indicated that seven autocovariances (i.e., correlation between error variances of an item measured at both time points) between dedication items 1, absorption items 1, vigor items 3, dedication items 4, vigor items 4, absorption items 6, and vigor items 6 needed to be estimated. Because of the large sample size, several autocovariances were statistically significant; hence autocovariances smaller than 0.10 were not estimated in the model, as they were considered to be of no practical importance.

Table 6 Time-invariance tests for the correlated three-factor model of the UWES-17 and the UWES-9

Alternative stability models	χ^2	df	$\Delta\chi^2(\Delta df)$	p	RMSEA	CFI	NNFI
1. Baseline stability model							
UWES-17	4207.72	518			0.060	0.95	0.94
UWES-17[a]	2901.24[a]	511			0.048[a]	0.97[a]	0.96[a]
UWES-9	1176.00	126			0.061	0.97	0.96
UWES-9[b]	816.29[b]	123			0.051[b]	0.98[b]	0.97[b]
2. Constrained stability model							
UWES-17[a]	3377.38[a]	525	2 vs. 1[a]		0.052[a]	0.96[a]	0.96[a]
			476.14 (14)	< 0.001			
UWES-17[c]	2915.97[c]	519	2 vs. 1[c]		0.048[c]	0.97[c]	0.96[c]
			14.73 (8)	= 0.065			
UWES-9[b]	992.68[b]	129	2 vs. 1[b]		0.055[b]	0.97[b]	0.97[b]
			176.39 (6)	< 0.001			
UWES-9[d]	830.32[d]	127	2 vs. 1[d]		0.050[d]	0.98[d]	0.97[d]
			14.03 (4)	= 0.007			

Baseline stability model = thresholds of observed variables set to be equal across two measurement points

Constrained stability model = thresholds of observed variables and factor loadings set to be equal across two measurement points

[a] Seven autocovariances of the UWES-17 estimated in the model

[b] Three autocovariances of the UWES-9 estimated in the model

[c] Seven autocovariances of the UWES-17 estimated in the model and the factor loadings of vigor items 2 and 4, dedication items 2 and 3, and absorption items 2 and 6 of the UWES-17 estimated as unequal across time

[d] Three autocovariances of the UWES-9 estimated in the model and the factor loadings of vigor item 2 and dedication item 2 of the UWES-9 estimated as unequal across time

After estimating these seven autocovariances, the fit of the baseline stability model clearly improved. Next, the equality of the factor loadings was tested. However, according to the chi-square difference test, the factor loadings equality assumption was not supported, as the constrained stability model yielded a significant loss of fit when compared to the baseline stability model (see Table 6). Thus, the size of the factor loadings did not remain unchanged over time, and the factor loadings could not be constrained to be equal but needed to be estimated freely. According to the modification indices, six factor loadings needed to be estimated freely. After the loadings of vigor items 2 and 4, dedication items 2 and 3, and absorption items 2 and 6 were estimated as unequal, the chi-square difference test did not produce a significant loss of fit. Since the six factor loadings were estimated as unequal, the UWES-17 did not measure work engagement in the same way at the two time points, and the thus time-invariance hypothesis (H3) regarding the UWES-17 was not supported.

As Table 6 also shows, the baseline stability model of the UWES-9 fitted the data well, but the modification indices indicated that three significant autocovariances between vigor items 3, dedication items 4, and absorption items 4 should be estimated (as previously, the cut-off point was 0.10). After estimating these autocovariances, the fit of the baseline stability model clearly improved. Next, the equality of the factor loadings was tested. However, the chi-square difference test did not support the factor loadings equality assumption, as the loss of fit for the constrained stability model was significant when
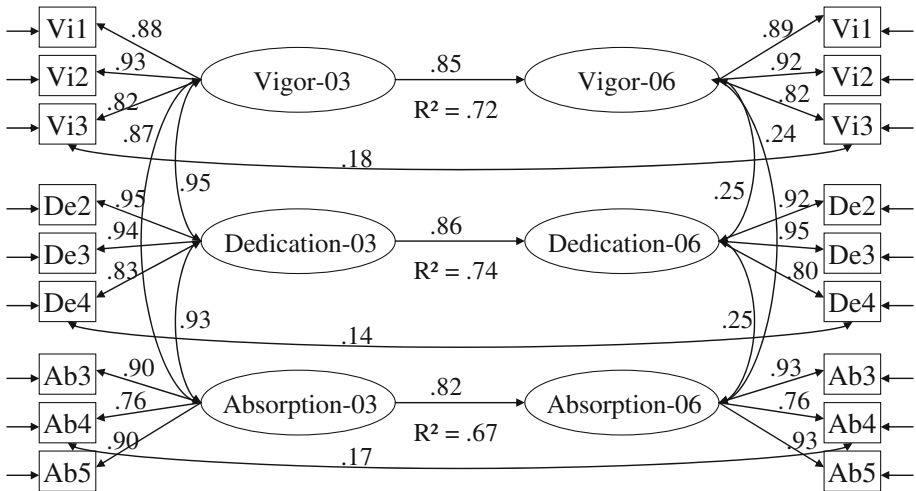
Fig. 2 Completely standardized solution for the final stability model of the correlated three-factor structure of the UWES-9. Three autocovariances between vigor items 3, dedication items 4, and absorption items 4 were estimated in the model. Factor loadings were constrained to be equal over time, with the exception that vigor item 2 and dedication item 2 were estimated as unequal over time

compared to the baseline stability model (see Table 6). Thus, the factor loadings did not remain equal over time, which impaired the fit of the constrained model. According to the modification indices, the factor loadings of vigor item 2 and dedication item 2 needed to be estimated as unequal. After these modifications, the factor loadings invariance assumption was supported, according to the chi-square difference test, at the 0.001 significance level. The final stability model of the UWES-9 is shown in Fig. 2. However, the detailed analysis (on the basis of the baseline stability model, completely standardized solution) revealed only minor, and for all practical purposes meaningless, variation in the factor loadings across time (between 0.01 and 0.04). Therefore, owing to the fact that the unequal factor loadings were unimportant and, that except for the chi-square difference test, the loss of fit was minor, the UWES-9 measured work engagement rather similarly over time. Thus the time-invariance hypothesis (H3) regarding the UWES-9 was relatively well supported.

Finally, the rank-order stabilities of work engagement factors were evaluated. However, as the structure of the UWES-17 did not remain invariant over time, the stability of work engagement was examined for the UWES-9. The standardized stability coefficients of the work engagement factors were high: vigor 0.85, dedication 0.86, and absorption 0.82 (see Fig. 2) and the stability hypothesis (H4) was fully supported. Thus, the reports of vigor, dedication and absorption remained highly stable over the three-year time-period and the proportion of the variance of the second measurement time explained by the first measurement time for each aspect of work engagement varied between 67% and 74%.

4 Discussion

Work engagement is one important construct of adults' happiness and well-being at work. The present study produced new knowledge about its measurement and stability. The study focused specifically on investigating the factor structure and factorial group and time invariance of the Finnish translations of the UWES-17 and the UWES-9 as well as the

rank-order stability of work engagement by using data from five different studies, one of which was a longitudinal study with a three-year follow-up time.

4.1 Work Engagement Consists of Three Interrelated Factors

The first hypothesis that work engagement consists of three correlated factors—vigor, dedication and absorption—was fully supported. As in previous CFA studies on work engagement (Hakanen 2002; Hallberg and Schaufeli 2006; Schaufeli and Bakker 2003; Schaufeli et al. 2002b, 2006), the theoretically based correlated three-factor structure of the UWES-17 and the UWES-9 showed good fit both for each sample and at both measurement points, as well as better fit with all the data sets than the alternative one-factor structure (H1). Also, as hypothesized, the three factors correlated highly.

However, the correlated three-factor structure did not show a flawless approximation with the data. In line with the previous studies, the loadings of absorption item 6 (“*It is difficult to detach myself from my job*”) and vigor item 6 (“*At my work I always persevere, even when things do not go well*”) of the UWES-17 were relatively low in one/two of the five samples. Moreover, the theoretically based structure of the UWES-9 also showed weakness. This shortcoming, however, seems to be related to one error covariance between absorption items 4 (“*I am immersed in my work*”) and 5 (“*I get carried away when I'm working*”), as the model fit clearly improved after allowing the error variances of these items to be correlated. This finding indicates that these items share some combined variance which the absorption factor cannot explain. As this error covariance appeared in all of the five samples, it seems that these two items overlap and to some extent measure the same thing. However, future studies need to replicate this finding before further conclusions can be drawn.

In line with the previous CFA studies, the results of this study reveal that work engagement can be considered both as a one-dimensional and as a three-dimensional construct, depending on the research purpose. High correlations between the three factors (from 0.83 to 0.97) would indicate a one-dimensional structure, but the better fit with the data of the correlated three-factor structure supports the three different, though highly correlated dimensions. Therefore, if the purpose is to study work engagement in general, a combined one-dimensional variable may be used, and if the purpose is to study the factors of work engagement, three separate dimensions may be used. However, from a practical viewpoint, the high correlations between the three factors indicate substantial overlap between them, and thus restrict their use as separate dimensions. Therefore, it seems to be reasonable to use the three factors separately only when conducting CFA and SEM analyses. The proposed solution is, however, largely a pragmatic one, which leaves unresolved the ultimate question of the one- vs. three-dimensionality of work engagement. In future research, it would, therefore, be worthwhile to pay attention to the criterion validity of work engagement.

4.2 The UWES-9 Measures Work Engagement similarly among Different Occupations and Over Time

The second hypothesis that the correlated three-factor structure of the UWES-17 and the UWES-9 would show factorial group invariance was only partially supported. In line with the results of the 14-item and 15-item versions of the UWES (see Schaufeli et al. 2002a; Xanthopoulou et al., in press), the group-invariance assumption of the correlated three-factor structure of the UWES-17 was not supported (H2). Specifically, while the factor

structure of the UWES-17 was similar across the occupational groups, the size of the factor loadings differed between the groups. Thus, the UWES-17 did not measure work engagement similarly among different occupations. As a matter of fact we cannot be sure what construct was measured by the UWES-17. However, in contrast to the previous group-invariance study of the UWES-9 (Schaufeli et al. 2006), the group-invariance assumption of the correlated three-factor structure of the UWES-9 was relatively well supported (H2). The structure of the UWES-9 remained largely the same across the five samples, which means that participants with different occupations interpreted the scale in a conceptually similar manner. Thus, the UWES-9 seems to have good construct validity.

The third hypothesis positing factorial time invariance for the correlated three-factor structure of the UWES-17 and the UWES-9 was also only partially supported. The factor structure of the UWES-17 was similar over time but the size of the factor loadings differed between the two time points. Thus, the UWES-17 did not measure work engagement similarly over time and the time-invariance assumption of the correlated three-factor structure of the UWES-17 was not supported (H3). Moreover, the structure of the UWES-9 did not fully meet the criteria of factorial time invariance either, as the size of the factor loadings differed over time. However, the differences were very small, varying from 0.01 to 0.04, and thus, for all practical purposes, meaningless. It was, therefore, concluded that the correlated three-factor structure of the UWES-9 was relatively time-invariant (H3) and the UWES-9 measured work engagement rather similarly over the three-year time-period. This also indicates that the short scale has good construct validity.

In all, both versions of the UWES verified the theory of work engagement as three dimensional; however, contrary to the UWES-17, the correlated three-factor structure of the UWES-9 remained relatively unchanged across both samples and time. Although the main purpose of the study was not to compare these two versions of the UWES, the results show, in line with the previous study (Schaufeli et al. 2006), that the use of the 9-item version of the scale can be recommended. The shorter version of the scale is also preferable for practical reasons—to reduce the likelihood of attrition a scale measuring a particular construct should have as few items as possible while remaining reliable and valid. However, more analyses on the psychometric qualities of the UWES-9 are needed, particularly on the structural invariance of the scale.

4.3 Work engagement is a Highly Stable Construct

The fourth hypothesis, that the rank-order stability of work engagement would be relatively high, was fully supported. However, as the structure of the UWES-17 did not remain unchanged over time, the stability of work engagement was studied only for the UWES-9. The stability of work engagement was very high (standardized stability coefficients varied between 0.82 and 0.86) and thus, as theoretically expected, the feelings at work engagement tended to be highly stable and long-lasting over the three-year follow-up (H4). The stability of work engagement is considered to be similar to that of its negative opposite—burnout (for the stability of burnout, see Schaufeli and Enzmann 1998). This assumption has also received support in previous longitudinal studies on work engagement (Llorens et al. 2007; Mauno et al. 2007; Schaufeli et al. 2006). In the present study, however, the stability coefficients of work engagement turned out to be somewhat higher, probably owing to the use of SEM, which yields error-free stability coefficients.

According to the definition, work engagement reflect on employee's present, though persistent and pervasive, state of mind and not a personality trait, which is a durable disposition reflecting a person's typical reaction (Schaufeli and Salanova 2007; Schaufeli

et al. 2002b). The high stability coefficients found for work engagement in this study raise, however, an important further question: how much influence do personal factors have on work engagement? The role of personality in work engagement is not yet well understood, although a recent study (see Langelaan et al. 2006) revealed that work engagement was related to high levels of extraversion and low levels of neuroticism. However, more studies on work engagement and personality are needed before further conclusions can be drawn. It is also important to bear in mind the context-specificity of the data. The longitudinal data set was restricted to Finnish dentists and was female-dominated. The respondents were highly educated, most had a permanent job contract and a long career history as a dentist and they experienced work engagement rather often (see Table 1).

In order to generalize the findings of this study, the stability of work engagement needs, therefore, to be investigated in a variety of other (unstable) contexts as well. It would also be interesting to study the stability of work engagement over a longer period with several measurements in order to investigate its stability in the long term. A longer follow-up with several measurement points would also allow investigation of the developmental trajectories of work engagement; utilizing a person-oriented approach would yield a more specific understanding of stability/change in work engagement than the conventional methods of the variable-centered approach (see, e.g., Laursen and Hoff 2006; Magnusson 1998).

4.4 Study Limitations

The present study has some limitations which need to be acknowledged. The main limitation is that the data consisted only of Finnish employees, who were mainly highly educated white-collar workers. Thus, the limitations on the sample restrict generalization of the findings and the results can be best generalized to Finnish white-collar workers. Since the UWES is expected to measure work engagement despite occupation or professional field, and since the structure of the UWES was not flawless even across the present white-collar occupations, the psychometric properties of the Finnish translation of the scale need also to be studied among blue-collar workers. The second limitation is that sample attrition analyses were not possible for all the samples. Therefore, it is possible that the samples obtained in the original studies were to some extent selective and so not representative with respect to those occupational groups. Finally, with respect to the structure of the UWES, one limitation is that the theoretically based correlated three-factor structure was not flawless in either version of the scale. However, overall the structure of the UWES-9 was good. On the other hand, the use of a multi-sample design, longitudinal data and sophisticated statistical methods can be considered the special strengths of this study. It should also be noted that this study was the first to examine the error-free rank-order stability of work engagement.

5 Conclusions

Happiness in life comes by many routes. Seen in this light, developing one's strengths and virtues in the major realms of living—of which work is one—becomes an important task. Work engagement with its positive outcomes both in and outside work (e.g., Schaufeli and Salanova 2007) seems to be among this desired task. Furthermore, since research on work-related happiness and positive well-being as well as the positive approach as a whole in the field of occupational health psychology is still rather new, few reliable measures have as

yet been developed to evaluate the positive pole. The UWES-9 seems to be a sound measure of work engagement and the use of the 9-item version of the scale can be recommended in future research on occupational well-being.

Appendix

Work and Well-being Survey (UWES) ©

The following 17 statements are about how you feel at work. Please read each statement carefully and decide if you ever feel this way about your job. If you have never had this feeling, write "0" (zero) in the space preceding the statement. If you have had this feeling, indicate how often you feel it by writing the number (from 1 to 6) that best describes how frequently you feel that way.

	Almost never	Rarely	Sometimes	Often	Very often	Always
0	1	2	3	4	5	6
Never	A few times a year or less	Once a month or less	A few times a month	Once a week	A few times a week	Every day

1. At my work, I feel that I am bursting with energy (VI1)*

2. I find the work that I do full of meaning and purpose (DE1)

3. Time flies when I'm working (AB1)

4. At my job, I feel strong and vigorous (VI2)*

5. I am enthusiastic about my job (DE2)*

6. When I am working, I forget everything else around me (AB2)

7. My job inspires me (DE3)*

8. When I get up in the morning, I feel like going to work (VI3)*

9. I feel happy when I am working intensely (AB3)*

10. I am proud of the work that I do (DE4)*

11. I am immersed in my work (AB4)*

12. I can continue working for very long periods at a time (VI4)

13. To me, my job is challenging (DE5)

14. I get carried away when I'm working (AB5)*

15. At my job, I am very resilient, mentally (VI5)

16. It is difficult to detach myself from my job (AB6)

17. At my work I always persevere, even when things do not go well (VI6)

* Shortened version (UWES-9); VI = Vigor; DE = Dedication; AB = Absorption

References

Bakker, A. B., Hakanen, J. J., Demerouti, E., & Xanthopoulou, D. (2007). Job resources boost work engagement, particularly when job demands are high. *Journal of Educational Psychology, 99*, 274–284. doi:10.1037/0022-0663.99.2.274.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606. doi:10.1037/0033-2909.88.3.588.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons, Inc.

Bollen, K. A., & Long, J. S. (1993). Introduction. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 1–9). Newbury Park, CA: Sage, cop.

Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling, 7*, 461–483. doi:10.1207/S15328007SEM0703_6.

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). Newbury Park, CA: Sage, cop.

Gray, E. K., & Watson, D. (2001). Emotion, mood, and temperament: Similarities, differences, and a synthesis. In R. L. Payne & C. L. Cooper (Eds.), *Emotions at work: Theory, research and applications for management* (pp. 21–43). Chichester: Wiley, cop.

Hakanen, J. (2002). Työuupumuksesta työn imuun—positiivisen työhyvinvointikäsitteen arviointimenetelmän suomalaisen version validointi opetusalan organisaatiossa [From burnout to work engagement—validation of the Finnish version of an instrument for measuring work engagement (UWES) in an educational organization]. *Työ ja ihminen, 16*, 42–58.

Hakanen, J. J., Bakker, A. B., & Demerouti, E. (2005). How dentists cope with their job demands and stay engaged: The moderating role of job resources. *European Journal of Oral Sciences, 113*, 479–487. doi: 10.1111/j.1600-0722.2005.00250.x.

Hakanen, J. J., Bakker, A. B., & Schaufeli, W. B. (2006). Burnout and work engagement among teachers. *Journal of School Psychology, 43*, 495–513. doi:10.1016/j.jsp.2005.11.001.

Hallberg, U. E., & Schaufeli, W. B. (2006). "Same same" but different? Can work engagement be discriminated from job involvement and organizational commitment? *European Psychologist,* 11, 119–127. doi:10.1027/1016-9040.11.2.119.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and application*. Thousand Oaks, CA: Sage, cop.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Jöreskog, K. G. (2005). *Structural equation modeling with ordinal variables using LISREL* (2nd ed.). Retrieved February 28, 2008, from http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf.

Jöreskog, K., & Sörbom, D. (1996a). *LISREL 8: User's reference guide*. Chicago: Scientific Software International, Inc.

Jöreskog, K., & Sörbom, D. (1996b). *PRELIS 2: User's reference guide*. Chicago: Scientific Software International, Inc.

Kinnunen, U., Feldt, T., & Mäkikangas, A. (2008). Testing the effort-reward imbalance model among Finnish managers: The role of perceived organizational support. *Journal of Occupational Health Psychology, 13*, 114–127.

Langelaan, S., Bakker, A. B., Van Doornen, L. J. P., & Schaufeli, W. B. (2006). Burnout and work engagement: Do individual differences make a difference? *Personality and Individual Differences, 40*, 521–532. doi:10.1016/j.paid.2005.07.009.

Laursen, B., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly, 52*, 377–389. doi:10.1353/mpq.2006.0029.

Llorens, S., Schaufeli, W., Bakker, A., & Salanova, M. (2007). Does a positive gain spiral of resources, efficacy beliefs and engagement exist? *Computers in Human Behavior, 23*, 825–841. doi: 10.1016/j.chb.2004.11.012.

Magnusson, D. (1998). The logic and implications of a person-oriented approach. In R. B. Cairns, L. R. Bergman, & J. Kagan (Eds.), *Methods and models for studying the individual: Essays in honour of Marian Radke-Yarrow* (pp. 33–64). Thousands Oaks, CA: Sage, cop.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *MBI: Maslach Burnout Inventory—Manual* (3rd Rev. ed.). Palo Alto, CA: Consulting Psychologists Press.

Maslach, C., & Leiter, M. P. (1997). *The truth about burnout: How organizations cause personal stress and what to do about it*. San Francisco: Jossey-Bass, Inc.

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology, 52*, 397–422. doi:10.1146/annurev.psych.52.1.397.

Mauno, S., Kinnunen, U., Mäkikangas, A., & Nätti, J. (2005). Psychological consequences of fixed-term employment and perceived job insecurity among health care staff. *European Journal of Work and Organizational Psychology, 14*, 209–237.

Mauno, S., Kinnunen, U., & Ruokolainen, M. (2007). Job demands and resources as antecedents of work engagement: A longitudinal study. *Journal of Vocational Behavior, 70*, 149–171. doi:10.1016/j.jvb.2006.09.002.

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika, 29*, 177–185. doi:10.1007/BF02289699.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543. doi:10.1007/BF02294825.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507–514. doi:10.1007/BF02296192.

Schaufeli, W. B. (2007a). *Literature list engagement*. Retrieved February 28, 2008, from http://www.Schaufeli.com.

Schaufeli, W. B. (2007b). *Tests*. Retrieved February, 28, 2008, from http://www.Schaufeli.com.

Schaufeli, W., & Bakker, A. (2003). *UWES Utrecht Work Engagement Scale. Preliminary Manual* [*Version 1, November 2003*]. Utrecht University: Occupational Health Psychology Unit.

Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire. A cross-national study. *Educational and Psychological Measurement, 66*, 701–716. doi:10.1177/0013164405282471.

Schaufeli, W., & Enzmann, D. (1998). *The burnout companion to study and practice: A critical analysis*. London: Taylor & Francis, cop.

Schaufeli, W. B., Martínez, I. M., Pinto, A. M., Salanova, M., & Bakker, A. B. (2002a). Burnout and engagement in university students: A cross-national study. *Journal of Cross-Cultural Psychology, 33*, 464–481. doi:10.1177/0022022102033005003.

Schaufeli, W., & Salanova, M. (2007). Work engagement: An emerging psychological concept and its implications for organizations. In S. W. Gilliland, D. D. Steiner, & D. P. Skarlicki (Eds.), *Research in social issues in management (Volume 5): Managing social and ethical issues in organizations* (pp. 135–177). Greenwich, CT: Information Age Publishers.

Schaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. B. (2002b). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies, 3*, 71–92. doi:10.1023/A:1015630930326.

Seligman, M. E. P. (2002). Positive psychology, positive prevention, and positive therapy. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 3–9). New York: Oxford University Press.

Seligman, M. E. P. (2003). Positive psychology: Fundamental assumptions. *The Psychologist, 16*, 126–127.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*, 5–14. doi:10.1037/0003-066X.55.1.5.

Turner, N., Barling, J., & Zacharatos, A. (2002). Positive psychology at work. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 715–728). New York: Oxford University Press.

Warr, P. (1990). The measurement of well-being and other aspects of mental health. *Journal of Occupational Psychology, 63*, 193–210.

Xanthopoulou, D., Bakker, A. B., Kantas, A., & Demerouti, E. (in press). The measurement of burnout and work engagement: A comparison of Greece and The Netherlands. *New Review of Social Psychology*.

Yu, C.–Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Retrieved February 28, 2008, from http://www.statmodel.com/download/Yudissertation.