Studia Psychologica, Vol. 68, No. 3, 2026, 256-279
https://doi.org/10.31577/sp.2026.03.948

Psychometric Evaluation of the Slovak Version of the 9-Item Utrecht Work Engagement Scale (UWES-9)

Ester Nosáľová, Simona Dudová
Department of Psychology, Faculty of Arts, Pavol Jozef Šafárik University in Košice, Košice, Slovak Republic

The objective was to evaluate the factorial and convergent validity, as well as the reliability of the 9-item Utrecht Work Engagement Scale (UWES-9) in the Slovak language. As a secondary objective, the psychometric properties of the ultra-short UWES-3 were examined. Study 1 included 532 employees (M_{age} = 46.2; 63.7% women) and Study 2 included 985 employees (M_{age} = 42.0; 48.5% women). Measures included the UWES-9, the BAT-12 (Burnout Assessment Tool), self-rated work performance, and the ISA-9 (Social, Intellectual, Affective Engagement Scale). Confirmatory factor analyses compared unidimensional and three-factor models. Global and local fit indices supported the three-factor structure in both studies. After incorporating two residual correlations identified in prior research, model fit further improved in Study 2. Internal consistency was high across studies and supported by average variance extracted, construct replicability, mean inter-item correlation, and item–total correlations. Test–retest (N = 353) analyses in Study 1 indicated good temporal stability. Measurement invariance across gender was supported in both studies, and across occupational groups (manual, non-manual, management) in Study 2. Convergent validity was demonstrated through negative associations with burnout and positive associations with work performance and ISA-9. Overall, the Slovak UWES-9 shows sound psychometric properties, with UWES-3 serving as a viable alternative when brevity is required.

Key words: work engagement, CFA, reliability, validity, burnout

Introduction

Work engagement is a widely recognized concept that plays a vital role in the work lives of everyone. It is not surprising that, as a result, many measurement methods have been developed. One of the most widely used is UWES – Utrecht Work Engagement Scale, which measures three dimensions of Work engagement – Vigor, Dedication, and Absorption. The existence of many revisions of the

Correspondence concerning this article should be addressed to Ester Nosáľová, Department of Psychology, Faculty of Arts, Pavol Jozef Šafárik University in Košice, Moyzesova 9, 040 01, Košice, Slovak Republic. E-mail: ester.nosalova@upjs.sk

Online supplement is availabble at

Received November 17, 2025

scale symbolizes the effort to keep the measurement short while still being an effective tool. The 9-item short version is one of the most recent revisions of the scale. This study aims to examine the psychometric properties of this version in the Slovak language.

Present Research

Engagement and Burnout

Having a positive attitude towards one's work and its components has a fundamental impact on work outcomes and satisfaction. Feeling energetic, connected to work, and the ability to perform work duties are core aspects of work-related attitudes and behavior reflected in work engagement (Schaufeli et al., 2006; Soane et al., 2012). The root of this state of mind is in seeing value and worth in the effort to complete daily tasks (Bakker et al., 2011).

Depending on the conceptual approach and model, there are various measurement tools. Most of the models have a strong foundation in theories revolving around motivation and needs. However, engagement was first introduced in close relation to burnout as two bipolar endpoints of an employee's well-being continuum, which can be affected by various circumstances (Nerstad et al., 2010). Therefore, engagement was viewed as the opposite of burnout, which was originally defined as a "a three-dimensional syndrome of emotional exhaustion (i.e., the draining of emotional resources because of demanding interpersonal contacts with others), depersonalization (i.e., a negative, callous, and cynical attitude towards the recipients of one's care or services), and a lack of personal accomplishment (i.e., the tendency to evaluate one's work with recipients negatively)". (Schaufeli et al., 2002, p. 72). This was later broadened to encompass all employees across various areas, exceeding helping professionals. Engagement was then viewed as the direct opposite of the burnout and its dimensions, and therefore characterized by energy, involvement and efficacy.

This view was later highly criticized, which led to the expansion of other approaches. In general, many definitions focus on the positivity and fulfillment that accompany a work-related state of mind. Nowadays, work engagement is often characterized by Vigor, Dedication, and Absorption (Schaufeli et al., 2006), which are still on the opposite side of burnout but should be viewed separately (not through the lens of burnout measurements). Moreover, longitudinal evidence indicates that work engagement is a relatively stable construct. Seppälä et al. (2009) report high test-retest correlations over a three-year period. This supports the notion that all three dimensions reflect enduring aspects of a positive work-related state of mind rather than transient moods.

Vigor signifies an employee's capacity to maintain elevated energy and vitality, even under stressful or complex work demands (Schaufeli et al., 2002; Schaufeli et al., 2006). Employees who exhibit Vigor approach tasks with determination rather than avoidance, reflecting what Bakker et al. (2011) call a "can-do attitude." This mindset facilitates continuous learning, as individuals remain open to mastering new tools, techniques, and problem-solving strategies.

Dedication reflects the employee's deep emotional commitment to their work and manifests strong feelings of value, enthusiasm, and personal challenge (Schaufeli et al., 2006). Rather than merely completing assigned duties, dedicated employees experience their roles as inherently meaningful and are driven by a purposeful connection to the broader objectives of their organization.

Absorption involves an intense state of focus in which individuals become wholly immersed in their work, often losing awareness of external distractions and of the passage of time (Schaufeli et al., 2002). This state closely aligns with Csikszentmihalyi's (1990) theory of flow, described as an optimal experience in which the individual is challenged in a way that matches their skill level, fostering deep engagement. Absorbed employees demonstrate remarkable concentration, frequently diving into tasks with such devotion that tasks feel inherently rewarding (Demerouti & Fullagar, 2016). In this scenario, employees often report heightened clarity, creativity, and efficiency, as they allocate their cognitive resources almost exclusively to problem-solving and task completion.

These components have become vital for understanding that fulfillment of the needs of organizations is largely dependent on the absorption, dedication, and energy of their employees (Nerstad et al., 2010). Therefore, the higher the level of the employees' work engagement, the better the results for their employers.

There are, of course, many other factors that play an important role in relation to work engagement, burnout, and work-related outcomes (e.g., autonomy, social support, performance feedback, personal resources, etc.). The complexity of relationships between these phenomena naturally led to the development of the most effective measurement tools.

Measurement Tools

As previously mentioned, engagement was measured by the same method as burnout at first. Originally, the Maslach Burnout Inventory (MBI; Maslach & Jackson, 1981) focused on three dimensions – emotional exhaustion, depersonalization, and personal accomplishment – and the resulting score showed the employee's tendency toward either burnout or engagement. Based on the broadening of the scope of the profession areas and criticism of certain aspects of the measurement, not only new revisions but also the development of different methods are expected in the coming years.

Schaufeli et al. (2020) introduced a new measure of burnout – the Burnout Assessment Tool (BAT) – that tries to acknowledge the MBI's conceptual, technical, and practical imperfections. BAT distinguishes between core symptoms (falling under the category of exhaustion, emotional impairment, cognitive impairment, and mental distance) and secondary/atypical symptoms (psychological distress, psychosomatic complaints, and depressed mood) (Schaufeli et al., 2020). A short work-related version of the BAT, containing 12 items focusing on core symptoms (3 items for each scale), was also developed and showed good internal consistency and validity (Schaufeli et al., 2020). Thus, the BAT, especially the short version, shows great potential as an effective tool for measuring burnout, with respect to four dimensions, but also coming to one underlying burnout factor.

The separation of burnout and engagement increased the need to develop a reliable measure of positive attitudes and behavior towards work. Until today, one of the most widely used measurement methods for work engagement is the Utrecht Work Engagement Scale (UWES), first introduced at the beginning of this century by Schaufeli and colleagues (2002) as a response to using MBI to examine engagement. They considered burnout and engagement to be opposite concepts that are very much intertwined but should be measured separately. Especially since some aspects are not direct opposites (e.g., reduced efficacy and absorption) but rather differ in their conceptualization.

The first version contained 17 items, comprising 3 scales (6 items for Vigor, 5 items for Dedication, and 6 items for Absorption), and did not load on one underlying general factor (Schaufeli et al., 2002; Schaufeli et al., 2006). A meta-analytic review by Christian and Slaughter (2007) indicated that UWES demonstrates a relatively stable three-factor structure compared to other engagement measures. Confirmatory studies across different occupational groups and countries have generally supported the three correlated dimensions, often demonstrating superior fit compared to one-factor alternatives (Nerstad et al., 2010; Seppälä et al., 2009).

Nevertheless, the dimensionality of work engagement remains a subject of an ongoing debate. Several studies have reported strong relationships between Vigor and Dedication, and occasionally among all three dimensions (e.g., Nerstad et al., 2010; Seppälä et al., 2009; Sinval et al., 2018). Seppälä et al. (2009) found that correlations among the factors were very high, which could justify treating engagement as a one-dimensional construct for some purposes, while the correlated three-factor solution generally provides a better fit in confirmatory analyses. Such findings have led some authors to test higher-order or bifactor models in which a general engagement factor coexists with specific dimensions. From a theoretical perspective, the three dimensions represent distinguishable components of a broader, positive work-related state of mind (Schaufeli et al., 2006). High correlations among them may therefore reflect their conceptual relatedness rather than a lack of multidimensionality.

Following these discussions, numerous studies have demonstrated satisfactory internal consistency of the UWES dimensions, with Cronbach's alphas ranging from .70 to .90, alongside acceptable factorial validity across various occupational and national contexts (Nerstad et al., 2010). Furthermore, consistent negative associations with burnout measures such as MBI and BAT provide additional support for the construct validity of engagement as a related yet distinct construct (Schaufeli et al., 2002; Schaufeli et al., 2020).

Later, a shorter 9-item version was developed, with three items dedicated to each scale (Schaufeli et al., 2006). The results of analyses of psychometric properties have been repeatedly acceptable (Balducci et al., 2010; Domínguez-Salas et al., 2022; Nerstad et al., 2010; Sinval et al., 2018), and some studies showed a more reasonable fit in the UWES-9 compared to the 17-item version (Nerstad et al., 2010). However, given the higher inter-factor correlations reported in several validation studies and the reduced number of items per dimension in the UWES-9, careful examination of its factorial structure in different linguistic and cultural contexts remains essential (Schaufeli et al., 2006; Nerstad et al., 2010).

The psychometric properties of UWES-9 have, to the best of our knowledge, never been investigated in Slovak. The purpose of this study is therefore to examine these properties and evaluate the suitability of the UWES-9 for use in the Slovak context.

Study 1

Method

Participants and Procedure

The sample consisted of 532 employed participants (339 women; 63.7%) recruited through a collection agency in October 2024. The mean age was 46.2 years (*SD* = 10.93). The majority of participants had completed secondary school education with a school-leaving exam (43.6%, *n* = 232) or a university with a Bachelor's/Master's degree (42.3%, *n* = 225). Some participants (8.6%, *n* = 46) reported secondary education without the school-leaving exam, a doctoral/third-degree qualification (4.9%, *n* = 26), or no/only primary education (0.6%, *n* = 3). All respondents provided informed consent, and their participation was anonymous and voluntary. Post-hoc power was calculated in WebPower based on the RMSEA close-fit test with df = 24/27, α = .05, H_0: RMSEA = .05, H_1: RMSEA = .08 was .92 and .95, respectively. To examine temporal stability, two additional waves of data collection were conducted. The second wave took place in January 2025 and included 416 employed participants from the original sample. The third wave was conducted in May 2025 and included 353 employed participants. These 353 participants completed all three waves and constituted the final sample used for the test–retest analyses.

Measures

Demographic variables included employment status, gender, and the highest level of completed education.

Work engagement was assessed using the 9-item *Utrecht Work Engagement Scale* (UWES-9; Schaufeli et al., 2006). The scale was translated into Slovak using the back-translation method involving four independent translators. It is comprised of nine items measuring three dimensions of work engagement: Vigor (e.g., *"At my work, I feel bursting with energy."*), Dedication (e.g., *"I am enthusiastic about my job."*), and Absorption (e.g., *"I feel happy when I am working intensely."*), with each dimension represented by three items rated on a 7-point scale. In addition, the ultra-short version of the UWES (UWES-3; Schaufeli et al., 2019), which includes one item per dimension, was also tested.

Burnout was assessed using the short 12-item version of the *Burnout Assessment Tool* (BAT; Schaufeli et al., 2020), which focuses on core symptoms of burnout. These include Exhaustion (e.g., *"I feel mentally exhausted."*), Mental distance (e.g., *"I struggle to find any enthusiasm for my work."*), Cognitive impairment (e.g., *"I have trouble staying focused."*), and Emotional impairment (e.g., *"I feel unable to control my emotions."*), with each dimension measured by three items. The scale demonstrated good internal consistency overall (McDonald's ω = .91), as well as for each subscale: Exhaustion (ω = .80), Mental distance (ω = .82), Cognitive impairment (ω = .85), and Emotional impairment (ω = .85).

Work performance was assessed with a single self-report item: "*How would you rate your current work performance from 0 to 100?* (higher numbers indicate better performance)."

Analyses

CFA. To evaluate the factorial validity of UWES-9, confirmatory factor analysis (CFA) using the *lavaan* package (Rosseel, 2012) in R was conducted. Three competing models were tested: 1) a unidimensional model with all items loading on a single Work Engagement factor; 2) a three-factor model consisting of Vigor, Dedication, and Absorption; and 3) a bifactor model with one general factor and three orthogonal group factors corresponding to Vigor, Dedication, and Absorption. However, the bifactor model exhibited a Heywood case, with item *uwes_7* showing an inadmissibly high standardized loading (> 9) and a large negative residual variance. Due to this empirical anomaly and concerns about over-factoring, results for the bifactor model are not reported further. For transparency, the factor loadings are available in the online supplementary materials for Study 1 (SM1) in Table S5.

All models were estimated using robust maximum likelihood estimation (MLR), with standardized latent variables. Model fit was evaluated using several fit indices. The chi-square goodness-of-fit test (χ^2) and its degrees of freedom (df) and *p*-value were reported. Complementary indices included the Comparative Fit Index (CFI) and the Tucker–Lewis Index (TLI), with values ≥ .90 indicating acceptable fit and ≥ .95 indicating excellent fit (Hu & Bentler, 1999; Kline, 2016). The Root Mean Square Error of Approximation (RMSEA) and its 90% confidence interval were also reported, with values ≤ .08 considered acceptable and ≤ .06 considered excellent (Hu & Bentler, 1999; Browne & Cudeck, 1993). The Standardized Root Mean Square Residual (SRMR) was interpreted using the conventional cutoff of ≤ .08 (Hu & Bentler, 1999). Additionally, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used for relative model comparisons, with lower values indicating better fit (Burnham & Anderson, 2004).

Local fit was evaluated by inspecting standardized residuals (acceptable threshold ±2.58), residual correlation matrices (acceptable threshold Δr < .10), and modification indices (MIs), with MIs greater than 10 flagged as potentially meaningful. These diagnostics were used to identify localized areas of misfit and guide interpretation, and no post-hoc model modifications were applied in Study 1. However, modification indices provided a basis for improving the models used in Study 2.

Invariance measurement of the UWES-9 across gender was examined using the multigroup confirmatory factor analysis (MG-CFA). Configural, metric, and scalar invariance models were estimated sequentially using robust maximum likelihood estimation (MLR). Changes in model fit were evaluated using differences in fit indices between nested models. Invariance was considered tenable when changes in fit indices met the recommended criteria (ΔCFI ≤ .010 and ΔRMSEA ≤ .015; Chen, 2007).

Discriminant validity. Due to the high intercorrelations between UWES factors, additional analyses were conducted to examine whether the three dimensions of work engagement are empirically distinguishable. Discriminant validity was assessed in two ways: 1) the fit of the unidimensional model was compared with the three-factor model using chi-square difference testing, with Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) also reported; 2) three alternative models were specified in which correlations between pairs of latent factors were constrained to 1 (Vigor–Dedication, Vigor–Absorption, and Dedication–Absorption). Each model fit of the constrained model was then compared with the original three-factor model using chi-square difference tests, and AIC and BIC values were examined for model comparison.

Reliability analyses. Reliability analyses were conducted for the full 9-item Utrecht Work Engagement Scale (UWES-9), its three subscales (Vigor, Dedication, and Absorption), as well as for the 3-item version (UWES-3). To evaluate internal consistency, Cronbach's alpha (α) and McDonald's omega total (ω) were calculated. In addition, average variance extracted (AVE), construct replicability (H), and mean inter-item correlation (MIC) were reported for a more detailed assessment of reliability. Corrected item-total correlations were also examined. Temporal stability (test–retest reliability) was assessed using Pearson correlations between scores obtained at the three measurement waves (T1, T2, and T3) among participants who completed all waves.

Validity. To assess convergent validity, we used Pearson correlations, as all variables were normally distributed based on skewness and kurtosis values. The overall BAT score, its four core symptom scales, and work performance were correlated with the overall UWES-9 score, the three UWES-9 subscales (Vigor, Dedication, Absorption), and the UWES-3.

Results

Confirmatory Factor Analysis

Item-level descriptive statistics for the UWES-9 items (means, standard deviations, medians, skewness, kurtosis, and response distributions) are provided in the Supplementary Materials for Study 1 (Table S1).

Table 1 presents the global fit indices for the unidimensional and three-factor models of the UWES-9. The three-factor model demonstrated superior fit, as indicated by lower χ^2, RMSEA, SRMR, AIC, and BIC values, and a higher CFI and TLI. Although both models showed generally acceptable fit overall, it is worth noting that the chi-square test was significant in both cases, suggesting some degree of model misfit, even though the statistic was substantially lower for the three-factor model. The RMSEA was higher than the acceptable range (.11) in the unidimensional model and within the acceptable range (≤ .08) for the three-factor model. However, the upper bound of the RMSEA confidence interval in the three-factor model (.10) approached the threshold for poor fit.

Local model fit was examined through standardized residuals and residual correlation matrices. In the unidimensional model, several standardized residual covariances exceeded the threshold of ±2.58. The largest residuals were observed between *uwes_2* and *uwes_1* (4.80), *uwes_8* and *uwes_9* (4.51), and *uwes_3* and *uwes_5* (2.30). This pattern was supported by the residual correlation matrix, in which some item pairs, e.g., *uwes_8* and *uwes_9,* showed higher residual correlations (r = .17). Complete local fit tables for the unidimensional model can be found in SM1 (Tables S6-S7).

In contrast, the three-factor model showed improved local fit, with fewer residuals exceeding critical values. Most residuals were closer to zero, though some discrepancies remained, e.g., between *uwes_3* and *uwes_5* (3.67), and between *uwes_8* and *uwes_9* (3.27). Overall, the residual correlations were smaller and more balanced across item pairs. The largest residual correlation again appeared between *uwes_8* and *uwes_9* (r = .08), which was considered acceptable. Complete local fit tables for the three-factor model can be found in SM1 (Tables S9-S10).

To support these findings and to identify potential improvements for Study 2, modification indices (MIs) were inspected. In the unidimensional model (see SM1; Table S8), the largest indices suggested potential improvements through correlated residuals between *uwes_1* and *uwes_2* (MI = 60.40), and

Table 1 *Comparison of model fit indices for the unidimensional and three-factor models of the UWES-9*

Model	χ^2 (df), p	CFI	TLI	RMSEA [90% CI]	SRMR	AIC	BIC
Unidimensional model	137.67 (27), p < .001	.96	.95	.11 [.09, .13]	.03	13655.61	13732.59
Three-factor model	77.40 (24), p < .001	.98	.97	.08 [.06, .10]	.02	13568.05	13657.86

Note. Robust (scaled) chi-square statistics are reported due to the use of the MLR estimator. CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion. The MLR estimator was used in all models.

between *uwes_8* and *uwes_9* (MI = 78.79). In the three-factor model, the MIs were generally lower (see SM1; Table S11), although some elevated values indicated possible cross-loadings (e.g., *Dedication to uwes_6*, MI = 43.89) and residual correlations between *uwes_1* and *uwes_2* (MI = 24.18), *uwes_3* and *uwes_5* (MI = 25.57), and *uwes_8* and *uwes_9* (MI = 45.27). However, because items *uwes_3* and *uwes_5* belong to different subscales, that residual correlation was not considered theoretically meaningful. Consequently, the same two correlated residuals (*uwes_1* and *uwes_2*, *uwes_8* and *uwes_9*) were selected for further testing in both the unidimensional and three-factor models in Study 2. To avoid capitalizing on sample-specific error and to maintain theoretical parsimony, these modifications were not applied to the current confirmatory model. Cross-loadings were also not introduced to preserve the theoretical structure of the UWES.

Table 2 presents standardized factor loadings and 95% confidence intervals for each item of the UWES-9 in both the unidimensional and three-factor models. Across both models, all items loaded significantly ($p < .001$) onto their respective latent factor(s), with loadings ranging from 0.70 to 0.90 in the unidimensional model, and from 0.75 to 0.91 in the three-factor model. Overall, factor loadings were consistently high across

Table 2 *Comparison of factor loadings for the unidimensional and three-factor models of the UWES-9*

Item	**Unidimensional model**		**Three-factor model**	
	Factor loadings	**CI**	**Factor loadings**	**CI**
UWES_1 (V)	0.82	[0.78, 0.85]	0.85	[0.81, 0.88]
UWES_2 (V)	0.86	[0.83, 0.89]	0.89	[0.86, 0.92]
UWES_3 (D)	0.90	[0.88, 0.92]	0.91	[0.89, 0.93]
UWES_4 (D)	0.88	[0.85, 0.90]	0.89	[0.86, 0.91]
UWES_5 (V)	0.83	[0.80, 0.86]	0.84	[0.81, 0.88]
UWES_6 (A)	0.80	[0.75, 0.84]	0.82	[0.77, 0.87]
UWES_7 (D)	0.82	[0.79, 0.86]	0.83	[0.80, 0.87]
UWES_8 (A)	0.77	[0.72, 0.82]	0.82	[0.78, 0.87]
UWES_9 (A)	0.70	[0.62, 0.76]	0.75	[0.68, 0.82]

Note. V = Vigor; D = Dedication; A = Absorption. All factor loadings were statistically significant at $p < .001$.

Table 3 *Correlations between latent factors in the three-factor model of UWES-9*

	Vigor	**Dedication**	**Absorption**
Vigor	-		
Dedication	.94	-	
Absorption	.86	.92	-

Note. All correlations were statistically significant at $p < .001$.

both models. Additional factor loadings statistics can be found in SM1 (Tables S3-S4). Table 3 shows the correlations between the latent factors of Vigor, Dedication, and Absorption in the three-factor model of the UWES-9. All correlations were statistically significant ($p < .001$) and notably high. The strongest correlation was observed between Vigor and Dedication ($r = .94$), followed closely by Dedication and Absorption ($r = .92$), and Vigor and Absorption ($r = .86$).

Invariance Measurement

Table 4 presents the measurement of invariance across gender. The configural model demonstrated good fit, constraining factor loadings to equality across groups did not meaningfully worsen model fit (ΔCFI = -.001; ΔRMSEA = -.003). Similarly, adding equality constraints on item intercepts resulted in negligible changes in fit indices (ΔCFI = -.001; ΔRMSEA = -.002).

Discriminant Validity

As shown in Table 5, model comparisons indicated that the three-factor model provided significantly better fit than the unidimensional model, $\Delta\chi^2$SB(3) = 48.65, $p < .001$. Similarly, all models in which correlations between pairs of latent factors were constrained to 1 showed significantly worse fit than the three-factor model (all $p < .001$), with higher AIC and BIC

Table 4 *Measurement invariance of the UWES-9 across gender*

Model	χ^2SB (df)	CFI	RMSEA	SRMR	Comparison	ΔCFI	ΔRMSEA
Configural	105.31 (48)	.979	.067	.026			
Metric	112.88 (54)	.978	.064	.035	M2 vs M1	-.001	-.003
Scalar	120.74 (60)	.978	.062	.036	M3 vs M2	-.001	-.002

Note. MLR = Maximum Likelihood Estimation with Robust (Satorra-Bentler) Standard Errors and Scaled Test Statistics. CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

Table 5 *Model comparisons testing discriminant validity of UWES dimensions*

Compared Models	$\Delta\chi^2$ SB (Δdf)	*p*	AIC (M1 / M2)	BIC (M1 / M2)
Three-factor vs. Unidimensional	48.65 (3)	< .001	13,568 / 13,656	13658 / 13733
Three-factor vs. Vigor–Dedication fixed to 1	17.11 (1)	< .001	13,568 / 13,607	13658 / 13692
Three-factor vs. Vigor–Absorption fixed to 1	33.43 (1)	< .001	13,568 / 13,638	13658 / 13723
Three-factor vs. Dedication–Absorption fixed to 1	19.38 (1)	< .001	13,568 / 13,607	13658 / 13692

Note. $\Delta\chi^2$ SB = Satorra–Bentler scaled chi-square difference test; Δdf = difference in degrees of freedom; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; M1 = three-factor model; M2 = comparison model.

values. Additionally, lavaan produced warnings indicating that the latent covariance matrix was not positive definite in all constrained models, further supporting the empirical distinguishability of the UWES dimensions.

Reliability Analysis

Although the UWES-3 was not included in the confirmatory factor analysis due to its shortness (three items), it was included in the reliability analysis. Table 6 summarizes reliability and internal consistency indices for the whole UWES-9, its subscales, and the ultra-short UWES-3 (for more detailed information, see SM1; Tables S12-S14). All scales demonstrated strong internal consistency, with Cronbach's α and McDonald's ω ranging from .84 to .95. The three-factor subscales showed acceptable to high AVE values (≥ .64). Construct replicability (H) values were also high (≥ .84), supporting the stability of the latent constructs. Mean inter-item correlations (MIC) and corrected item–total correlations fell within recommended ranges. The UWES-3 also demonstrated satisfactory internal consistency despite its brevity.

Table 7 presents the correlations between the UWES total scores and subscale scores across the three measurement waves. All correlations were positive and statistically significant ($p < .001$). Test–retest correlations for the UWES-9 ranged from $r = .77$ to .82. For the subscales, correlations ranged from $r = .71$ to .76 for Vigor, $r = .74$ to .81 for Dedication, and $r = .65$ to .75 for Absorption.

Validity Analysis

Table 8 presents Pearson correlations between work engagement, burnout indicators, and self-rated work performance. The UWES-9 total score showed the weakest correlations with burnout and its symptoms compared to the UWES-3 and the UWES-9 subscales. It was negatively correlated with the BAT total score ($r = -.25$) and all its subscales, with the weakest correlation observed for Emotional Impairment ($r = -.11$, $p = .01$). Among the UWES-9 subscales, Vigor and Dedication showed the strongest correlations with the BAT total score ($r = -.61$ and $r = -.55$, respectively), while Absorption had a somewhat weaker association ($r = -.40$). Vigor and Dedication were also most strongly negatively correlated with Mental Distance ($r = -.65$ and $r = -.64$, respectively), whereas Absorption showed generally weaker but still significant negative associations

Table 6 *Reliability and internal consistency indices for the UWES-9, subscales of UWES-9, and UWES-3*

Scale / Subscale	**α**	**ω**	**AVE**	**H**	**MIC**	**Item-Total r (range)**
UWES-9	.95	.95	.67	.95	.67	.71 – .89
Vigor	.89	.89	.74	.90	.74	.80 – .88
Dedication	.91	.91	.77	.92	.77	.83 – .87
Absorption	.84	.85	.64	.84	.64	.72 – .82
UWES-3	.86	.86	-	-	.67	.75 – .82

Note. α = Cronbach's alfa; ω = McDonald's omega; AVE = Average variance extracted; H = Construct replicability; MIC = Mean inter-item correlations. Item–total correlations are corrected for overlap.

across all burnout dimensions. The UWES-3 followed a similar pattern but with overall stronger correlations, including a negative correlation with BAT (r = -.55) and the weakest correlation with Emotional Impairment (r = -.34).

Mirroring the pattern seen with burnout (i.e., stronger associations at the subscale level than at the UWES-9 total), work performance correlated positively with all engagement indicators. The weakest association was with the UWES-9 total (r = .34), while

Table 7 *Means, standard deviations, and intercorrelations of the UWES total score and subscales across three waves*

	Scale / Subscale	M	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	UWES-9 ^{T1}	3.31	1.24	-											
(2)	UWES-9 ^{T2}	3.64	1.32	.77	-										
(3)	UWES-9 ^{T3}	3.58	1.33	.79	.82	-									
(4)	Vigor ^{T1}	2.99	1.30	.93	.71	.74	-								
(5)	Vigor ^{T2}	3.35	1.38	.71	.92	.76	.71	-							
(6)	Vigor ^{T3}	3.26	1.41	.74	.75	.94	.76	.75	-						
(7)	Dedication ^{T1}	3.49	1.43	.95	.73	.79	.84	.66	.71	-					
(8)	Dedication ^{T2}	3.81	1.48	.74	.95	.79	.67	.83	.70	.74	-				
(9)	Dedication ^{T3}	3.74	1.50	.79	.81	.96	.73	.74	.86	.81	.81	-			
(10)	Absorption ^{T1}	3.45	1.29	.90	.68	.67	.74	.59	.59	.78	.63	.64	-		
(11)	Absorption ^{T2}	3.77	1.36	.70	.92	.75	.61	.77	.65	.65	.83	.72	.70	-	
(12)	Absorption ^{T3}	3.75	1.33	.69	.74	.92	.59	.64	.76	.68	.69	.82	.65	.75	-

Note. All correlations were statistically significant at p < .001.

Table 8 *Pearson correlations between Work Engagement Measures (UWES-9, UWES-3, Subscales) and Burnout Indicators (BAT and Subscales)*

	Scale / Subscale	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	UWES-9	-										
(2)	UWES-3	.60	-									
(3)	Vigor	.61	.83	-								
(4)	Absorption	.88	.76	.77	-							
(5)	Dedication	.64	.92	.85	.80	-						
(6)	BAT	-.25	-.55	-.61	-.40	-.55	-					
(7)	Exhaustion	-.18	-.46	-.57	-.31	-.45	.83	-				
(8)	Mental Distance	-.35	-.62	-.65	-.50	-.64	.85	.64	-			
(9)	Cognitive Impairment	-.17	-.38	-.45	-.29	-.39	.82	.56	.59	-		
(10)	Emotional Impairment	-.11	-.34	-.34	-.22	-.33	.79	.49	.55	.57	-	
(11)	Work Performance	.34	.41	.45	.43	.45	-.44	-.32	-.44	-.44	-.26	-

Note. All correlations are significant at p < .001, except for the correlation between UWES-9 and Emotional Impairment (r = -.11), which was significant at p = .01. UWES = Utrecht Work Engagement Scale; BAT = Burnout Assessment Tool.

the UWES-9 subscales showed moderately stronger relations: Vigor (r = .45), Absorption (r = .43), and Dedication (r = .45). The UWES-3 total yielded a similar, moderately positive correlation (r = .41).

Study 2

Participants and Procedure

After excluding participants who failed the attention checks (n = 145), the final sample consisted of 985 employed participants recruited through a data collection agency in June 2025. Of these, 507 were men (51.5%), and 478 were women (48.5%). Participants' ages ranged from 20 to 65 years, with a mean of 42 years (SD = 10.9). The majority of respondents were employed full time (86.4%, n = 851), while some participants reported self-employment or freelance work (8.3%, n = 82), part-time employment (3.1%, n = 31), or business ownership or executive roles (2.1%, n = 21). Participants were classified into three occupational groups: manual workers (28.9%, n = 285), non-manual employees (44.6%, n = 439), and management (26.5%, n = 261). On average, participants had been in their current employment for 9.7 years (SD = 8.9) and reported working 8.2 hours per day (SD = 2.5). All participants provided informed consent, and their participation was anonymous and voluntary. Post-hoc RMSEA power (WebPower; df = 25/22; α = .05; H_0 = .05, H_1 = .08) exceeded .99 for both factor solutions.

Measures

Demographic variables. In addition to the variables measured in Study 1, employment status was recorded in more detail, and information about the length of current employment and average working hours per day was also included.

Work engagement was measured using the 9-item *Utrecht Work Engagement Scale* (UWES-9; Schaufeli et al., 2006), identical to the instrument used in Study 1 (see *Measures*, Study 1). Additionally, work engagement was also measured using the *Intellectual, Social, and Affective Engagement Scale* (ISA-9; Soane et al., 2012). The scale consists of nine items capturing three dimensions of engagement: intellectual (e.g., *"I focus hard on my work"*), social (e.g., *"I share the same work values as my colleagues"*), and affective (e.g., *"I feel positive about my work"*), rated on a 7-point scale. The scale showed good internal consistency overall (McDonald's ω = .89), as well as for each subscale: Intellectual (ω = .87), Social (ω = .87), and Affective (ω = .90) engagement.

Analyses

CFA. To provide additional support for the tested models, confirmatory factor analyses incorporating the modification indices identified in Study 1 were conducted. Two models were tested: 1) a unidimensional model with all items loading on a single Work Engagement factor, and 2) a three-factor model consisting of Vigor, Dedication, and Absorption. In both models, the following residual correlations were included: a) between *uwes_1* and *uwes_2*, and b) between *uwes_8* and *uwes_9*. As mentioned in Study 1, cross-loadings were not introduced to maintain theoretical parsimony, and the residual correlation between *uwes_3* and *uwes_5* was also excluded because these items belong to different subscales.

To ensure comparability with Study 1, all models were estimated using robust maximum likelihood (MLR) and evaluated using the same set of fit indices (χ^2, CFI, TLI, RMSEA, SRMR, AIC, BIC) and the same evaluation criteria. Local fit was also assessed by inspecting standardized residuals, residual correlation matrices, and modification indices (MIs).

Invariance measurement of the UWES-9 across gender was examined in the same way as in Study 1. In Study 2, we also examined invariance across occupational level, which was by grouping six job categories into three broader categories: manual workers/blue collar (employee without formal qualification and qualified worker or foreman/team leader), non-manual employees/white collar (clerical or administrative employee and medium-qualified employee), and management (highly qualified employee, middle management/executive and senior manager or top-level executive), following the same criteria specified for gender in Study 1.

Discriminant validity. In Study 2, discriminant validity was assessed using the same procedure as in Study 1. Specifically, the fit of the unidimensional model was compared with that of the three-factor model, and additional models with correlations between pairs of latent factors constrained to 1 were compared with the three-factor solution. The only difference was that the model modifications described in the previous section were included in all estimated models.

Reliability analyses. The same reliability analyses as in Study 1 were conducted for the UWES-9, its three subscales (Vigor, Dedication, Absorption), and the UWES-3. Internal consistency was assessed using Cronbach's alpha (α) and McDonald's omega total (ω), along with additional indices of reliability: AVE, construct replicability (H), and mean inter-item correlation (MIC).

Validity analyses. To assess convergent validity, we used Pearson correlations to assess whether the overall ISA (Intellectual, Social, Affective Work Engagement Scale) score and its three subscales correlated with the overall UWES-9 score, the three UWES-9 subscales (Vigor, Dedication, Absorption), and the UWES-3.

Results

Confirmatory Factor Analysis

Item-level descriptive statistics for the UWES-9 items (means, standard deviations, medians, skewness, kurtosis, and response distributions) are provided in the online SM2 (Table S1).

Table 9 presents the global fit indices for the unidimensional and three-factor models of the UWES-9. Overall, the unidimensional model demonstrated a slightly better fit, with lower χ^2, RMSEA, AIC, and BIC values, and a higher TLI value. The CFI and TLI were identical or nearly identical across both models. Except for the chi-square test, which was significant in both cases, all fit indices indicated a very good model fit.

Table 9 *Comparison of model fit indices for the unidimensional and three-factor models of the UWES-9*

Model	χ^2 (df), p	CFI	TLI	RMSEA [90% CI]	SRMR	AIC	BIC
Unidimensional model	79.84 (25), *p* <.001	.99	.98	.06 [.05, .07]	.02	25260.28	25358.14
Three-factor model	57.95 (22), *p* <.001	.99	.99	.05 [.04, .07]	.02	25231.48	25344.01

Note. Robust (scaled) chi-square statistics are reported due to the use of the MLR estimator. CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion. The MLR estimator was used in all models.

Local model fit was examined through standardized residuals and residual correlation matrices. In the unidimensional model, some residuals exceeded the acceptable threshold (±2.58), with the largest values observed between *uwes_7* and *uwes_5* (-5.25), *uwes_5* and *uwes_2* (2.87), and *uwes_8* and *uwes_6* (2.84). On the other hand, the residual correlation matrix showed acceptable local fit (Δr < .10) with the highest being between *uwes_8* and *uwes_6* (r = .06). Complete local fit tables for the unidimensional model can be found in the online supplementary materials for Study 2 (SM2) in Tables S5-S6.

Local model fit for the three-factor model was slightly better than that of the unidimensional model. The largest residuals were observed between *uwes_7* and *uwes_5* (-4.25), *uwes_8* and *uwes_5* (-2.78), and *uwes_4* and *uwes_3* (-2.65). The residual correlation matrix showed only small and acceptable residual correlations, the highest being between *uwes_8* and *uwes_1* (r = .04), and between *uwes_9* and *uwes_7* (r = .04). Complete local fit tables for the unidimensional model can be found in SM2 (Tables S8-S9).

Modification indices (MIs) were examined to evaluate potential areas for model improvement, although no further modifications were applied in Study 2. In the unidimensional model (see SM2; Table S7), the largest MIs suggested possible correlated residuals between *uwes_5* and *uwes_7* (MI = 30.96), *uwes_6* and *uwes_8* (MI = 21.14), and *uwes_2* and *uwes_5* (MI = 16.08). In the three-factor model (see SM2; Table S10), higher MIs indicated potential cross-loadings (e.g., *Vigor to uwes_7*, MI = 19.43; *Absorption to uwes_1*, MI = 14.30), and correlated residuals between *uwes_5* and *uwes_7* (MI = 21.02), and between *uwes_1* and *uwes_8* (MI = 12.87). However, because the goal of Study 2 was to confirm the models specified a priori based on Study 1, these additional modifications were not implemented.

Table 10 presents standardized factor loadings and 95% confidence intervals for each item of the UWES-9 in both the unidimensional and three-factor models. Across both models, all items loaded significantly (p < .001) onto their latent factor(s), with loadings ranging from 0.56 to 0.91 in the unidi-

Table 10 *Comparison of factor loadings for the unidimensional and three-factor models of the UWES-9*

Item	Unidimensional model		Three-factor model	
	Factor loadings	**CI**	**Factor loadings**	**CI**
UWES_1 (V)	0.78	[0.75, 0.81]	0.79	[0.76, 0.82]
UWES_2 (V)	0.82	[0.79, 0.85]	0.84	[0.81, 0.87]
UWES_3 (D)	0.91	[0.89, 0.92]	0.91	[0.89, 0.93]
UWES_4 (D)	0.88	[0.86, 0.90]	0.89	[0.87, 0.91]
UWES_5 (V)	0.86	[0.84, 0.88]	0.87	[0.85, 0.90]
UWES_6 (A)	0.79	[0.76, 0.83]	0.83	[0.79, 0.86]
UWES_7 (D)	0.83	[0.80, 0.86]	0.84	[0.91, 0.86]
UWES_8 (A)	0.69	[0.65, 0.74]	0.72	[0.68, 0.77]
UWES_9 (A)	0.56	[0.50, 0.62]	0.57	[0.50, 0.64]

Note. V = Vigor; D = Dedication; A = Absorption. All factor loadings were statistically significant at p < .001.

mensional model, and from 0.57 to 0.91 in the three-factor model. Additional factor loadings statistics can be found in SM2 (Tables S3-S4).

Table 11 presents the correlations between the latent factors of Vigor, Dedication, and Absorption in the three-factor model. All correlations were statistically significant ($p < .001$), positive, and strong. The strongest correlation was between Vigor and Dedication ($r = .97$), followed by Vigor and Absorption ($r = .95$), and Absorption and Dedication ($r = .94$).

Invariance Measurement

Measurement invariance of the UWES-9 across gender was also examined in Study 2 using multi-group CFA and is presented in Table 12. The configural model demonstrated acceptable fit. Metric invariance was supported by negligible changes in model fit (ΔCFI = -.001; ΔRMSEA = -.000) when constraining factor loadings to equality. Scalar invariance was also supported, when constraining item

Table 11 *Correlations between latent factors in the three-factor model of UWES-9*

	Vigor	**Dedication**	**Absorption**
Vigor	-		
Dedication	.97	-	
Absorption	.95	.94	-

Note. All correlations were statistically significant at $p < .001$.

Table 12 *Measurement invariance of the UWES-9 across gender*

Model	**χ^2SB (df)**	**CFI**	**RMSEA**	**SRMR**	**Comparison**	**ΔCFI**	**ΔRMSEA**
Configural	84.83 (44)	.990	.043	.017			
Metric	95.65 (50)	.989	.043	.031	M2 vs M1	-.001	-.000
Scalar	108.31 (56)	.988	.044	.032	M3 vs M2	-.002	.000

Note. MLR = Maximum Likelihood Estimation with Robust (Satorra-Bentler) Standard Errors and Scaled Test Statistics. CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

Table 13 *Measurement invariance of the UWES-9 across occupation*

Model	**χ^2SB (df)**	**CFI**	**RMSEA**	**SRMR**	**Comparison**	**ΔCFI**	**ΔRMSEA**
Configural	98.62 (66)	.992	.039	.020			
Metric	107.87 (78)	.993	.034	.026	M2 vs M1	.001	-.005
Scalar	129.61 (90)	.991	.037	.030	M3 vs M2	-.002	.002

Note. MLR = Maximum Likelihood Estimation with Robust (Satorra-Bentler) Standard Errors and Scaled Test Statistics. CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

intercepts likewise did not meaningfully worsen model fit (ΔCFI = -.002; ΔRMSEA = -.000).

As shown in Table 13, measurement of invariance across occupational levels yielded results similar to those for measurement across gender. The configural model showed acceptable fit. Constraining factor loadings (ΔCFI = .001; ΔRMSEA = -.005) and item intercepts (ΔCFI = -.002; ΔRMSEA = .002) resulted in negligible changes in model fit, supporting metric and scalar invariance across occupational groups in Study 2.

Discriminant Validity

As shown in Table 14, the three-factor model demonstrated significantly better fit than the unidimensional model, $\Delta\chi^2$SB(3) = 21.43, p < .001. Likewise, all models in which correlations between pairs of latent factors were constrained to 1 fit the data significantly worse than the three-factor model (all $p \leq .012$), as reflected in higher AIC and BIC values. Similar to Study 1, lavaan produced warnings indicating that the latent covariance matrix was not positive definite in all constrained models.

Table 14 *Model comparisons testing discriminant validity of UWES dimensions*

Compared Models	**$\Delta\chi^2$ SB (Δdf)**	***p***	**AIC (M1 / M2)**	**BIC (M1 / M2)**
Three-factor vs. Unidimensional	21.43 (3)	< .001	25,232 / 25,260	25344 / 25358
Three-factor vs. Vigor–Dedication fixed to 1	11.85 (1)	< .001	25,232 / 25,248	25344 / 25355
Three-factor vs. Vigor–Absorption fixed to 1	6.26 (1)	.012	25,232 / 25,241	25344 / 25348
Three-factor vs. Dedication–Absorption fixed to 1	11.18 (1)	< .001	25,232 / 25,248	25344 / 25355

Note. $\Delta\chi^2$ SB = Satorra–Bentler scaled chi-square difference test; Δdf = difference in degrees of freedom; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; M1 = three-factor model; M2 = comparison model.

Table 15 *Reliability and internal consistency indices for the UWES-9, subscales of UWES-9, and UWES-3*

Scale / Subscale	**α**	**ω**	**AVE**	**H**	**MIC**	**Item-Total r (range)**
UWES-9	.94	.93	.64	.94	.64	.59 – .89
Vigor	.91	.91	.77	.87	.73	.79 – .88
Dedication	.91	.91	.77	.91	.77	.84 – .87
Absorption	.78	.70	.50	.84	.75	.65 – .80
UWES-3	.84	.84	-	-	.64	.71 – .82

Note. α = Cronbach's alfa; ω = McDonald's omega; AVE = Average variance extracted; H = Construct replicability; MIC = Mean inter-item correlations. Item–total correlations are corrected for overlap.

Table 16 *Pearson correlations between UWES (UWES-9, UWES-3, subscales of UWES-9) and ISA (ISA-9, Subscales of ISA-9)*

	Scale / Subscale	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
(1)	UWES-9	-								
(2)	UWES-3	.95	-							
(3)	Vigor	.93	.90	-						
(4)	Absorption	.89	.86	.73	-					
(5)	Dedication	.95	.88	.84	.77	-				
(6)	ISA-9	.72	.70	.71	.58	.69	-			
(7)	Intellectual Engagement	.59	.59	.54	.55	.54	.80	-		
(8)	Social Engagement	.39	.38	.40	.29	.38	.79	.40	-	
(9)	Affective Engagement	.80	.76	.81	.61	.78	.86	.64	.47	-

Note. All correlations are significant at $p < .001$. UWES = Utrecht Work Engagement Scale; ISA-9 = Intellectual, Social, Affective Engagement Scale.

Reliability Analysis

Table 15 presents reliability and internal consistency indices for the whole UWES-9, its subscales, and the ultra-short UWES-3 (for more detailed information see SM2; Tables S11-S13). All scales demonstrated acceptable internal consistency, with Cronbach's α and McDonald's ω ranging from .70 to .94. Additionally, acceptable AVE values (≥ .50) as well as high values (≥ .84) of construct replicability (H) (≥ .84) were shown. Mean inter-item correlations (MIC) and corrected item–total correlations were also satisfactory.

Validity Analysis

As shown in Table 16, both the UWES-9 and UWES-3 demonstrated very similar correlation patterns (both in direction and effect size), with the strongest associations observed with Affective Engagement (r = .80 and .76, respectively), followed by the overall ISA-9 score (r = .72 and .70, respectively) and Intellectual Engagement (r = .59 for both), while the weakest relationships were found with Social Engagement (r = .39 and .38, respectively). The UWES subscales of Vigor and Dedication displayed the same trend, correlating most strongly with Affective Engagement (r = .81 and .78, respectively), then with the overall ISA-9 (r = .71 and .69, respectively) and Intellectual Engagement (r = .54 for both), and least with Social Engagement (r = .40 and .38, respectively). Absorption showed the weakest correlations overall, though it followed the same pattern. The strongest association was shown with Affective Engagement (r = .61), then with the overall ISA-9 score (r = .69), and weakest with Social Engagement (r = .29).

Discussion

The aim of Study 1 and Study 2 was to investigate and report on the factorial validity, invariance measurement, discriminant validity, reliability, and convergent validity of the short 9-item version of UWES as the main method for measuring work engagement and its three dimensions: Vigor, Dedication, and Absorption. In addition to evaluating the UWES-9, we also examined the psychometric performance of the ultra-short UWES-3 to determine whether this brief measure could serve as a reliable alternative in contexts where survey length is limited.

The primary aim of the confirmatory factor analyses was to compare the unidimensional and three-factor structures of the UWES-9 and to examine whether the model refinements identified in Study 1 (correlating error terms for *UWES_1* & *UWES_2,* and UWES-8 & UWES-9) would replicate in Study 2. These correlated residuals were retained not only on the basis of empirical support from Study 1 but also because the same modifications have been reported in previous validation research (Schaufeli et al., 2019). Across both studies, the three-factor model demonstrated superior global fit compared to the unidimensional model, as reflected in lower χ^2, RMSEA, SRMR, AIC, and BIC values, and higher CFI. Although the chi-square statistic was significant in all models, other fit indices reliably indicated that the three-factor structure provided a better representation of the data. Although the unidimensional structure in Study 1 yielded acceptable fit indices overall, its RMSEA exceeded acceptable thresholds, whereas the three-factor model showed RMSEA values within the recommended range. After implementing model modifications from Study 1 in Study 2, all global fit indices (except the chi-square test, which was significant in both models) indicated a good model fit, supporting the replicability of these adjustments. Local fit indices also favored the three-factor solution, with consistently higher standardized factor loadings than the unidimensional model in both studies.

From a more critical perspective, there are three main points worth mentioning. First, even though the chi-square test is often overlooked in otherwise well-fitting models and consistently significant in at least 12 countries (Domínguez-Salas et al., 2022; Schaufeli et al., 2006; Schaufeli et al., 2019; Zecca et al., 2015) and different versions of UWES (Carmona-Halty et al., 2019; Dimitriadou et al., 2020), it still suggests somewhat imperfect model fit. On the other hand, the chi-square test evaluates exact model fit, and larger samples increase the likelihood of detecting even small deviations between the model and the observed data (Kline, 2016; Marsh et al., 2004), which should be tentatively accepted or rejected and interpreted alongside other indices. Second, some discrepancies were observed among global fit indices, particularly between TLI/CFI/SRMR and RMSEA. While CFI and SRMR indicated good model fit, RMSEA values were elevated in the unidimensional model in Study 1. While CFI, TLI, and SRMR indicated that the models reproduced the observed covariance structure reasonably well, the elevated RMSEA values suggested remaining approximation error, indicating that the model fit was not optimal.

Third, although global fit indices were acceptable, inspection of local fit suggested several areas of localized strain. For example, some standardized residuals exceeded recommended thresholds in both studies, albeit to a lesser extent in the three-factor solution. On the other hand, residual correlations remained in the recommended thresholds with one exception in Study 1, which was resolved by implementing modifications. To our knowledge, most published UWES validation studies do not systematically report local fit diagnostics, making direct comparison challenging. Nonetheless, these results suggest that even when global fit supports the three-factor structure, certain item relations may reflect shared wording or content beyond the latent factors, warranting potential refinement.

Measurement invariance analyses indicated that the UWES-9 operates equivalently across men and women in both studies, as configural, metric, and scalar invariance were supported. In Study 2, the scale also demonstrated invariance across occupational levels (manual, non-manual, and management). These findings provide evidence that the UWES-9 measures work engagement consistently across demographic and occupational groups, supporting the comparability of scores and the use of the scale in diverse working populations.

Despite this empirical advantage of the three-factor structure, the latent factors of Vigor, Dedication, and Absorption were highly intercorrelated in both studies, indicating conceptual overlap as previously mentioned by other studies (Nerstad et al., 2010; Schaufeli et al., 2006). This led us to examine discriminant validity using two approaches. First, we compared the fit of the unidimensional and three-factor models. Second, we estimated models in which correlations between pairs of latent factors were constrained to 1. In both studies, the three-factor model demonstrated better fit than the unidimensional model and all constrained models. Moreover, the constrained models resulted in non-positive definite latent covariance matrices, indicating improper solutions when factor correlations were fixed to unity. Together, these findings suggest that, despite the high intercorrelations among the UWES dimensions, treating Vigor, Dedication, and Absorption as distinguishable yet closely related components of work engagement provides a more appropriate representation of the construct. This is in line with previous research (Nerstad et al., 2010; Seppälä et al., 2009). We agree with the notion that very high correlations could justify treating engagement as a one-dimensional construct for some purposes (Seppälä et al., 2009), but overall, they may reflect their close conceptual association rather than a lack of multidimensionality, as supported by discriminatory analyses.

Taken together, the findings from the confirmatory factor analysis are in line with the results of Schaufeli et al. (2006), who likewise reported a slightly better fit for the three-factor model, along with high intercorrelations among subscales across countries. This pattern has since been replicated in multiple validation studies across different cultural contexts and version variations of UWES (Balducci et al., 2010; Domínguez-Salas et al., 2022; Schaufeli et al., 2006; Schaufeli et al., 2019; Sinval et al., 2018; Zecca et al., 2015). Nevertheless, some authors have argued for a two-factor structure (Dimitriadou et al., 2020; Nerstad et al., 2010), which was not examined in the present studies.

The aim of the reliability analyses was to examine the internal consistency of the UWES-9 and its subscales, as well as the UWES-3. Across both studies, all versions demonstrated strong internal consistency. The UWES-9 showed high reliability in each sample, although slightly lower item–rest correlations emerged in Study 2. However, they remained within acceptable psychometric ranges. Each UWES-9 subscale (Vigor, Dedication, Absorption) also displayed strong reliability, with the Dedication dimension consistently yielding the highest coefficients.

Average Variance Extracted (AVE) values for all subscales indicated that an acceptable proportion of item variance was accounted for by their latent factors. Although the AVE for Absorption in Study 2 was somewhat lower than in Study 1, it still met recommended thresholds. Construct replicability (H) values were high across both studies, suggesting that the latent factors can be reliably replicated in future studies. Additionally, mean inter-item correlations (MIC) and total-rest correlations provided no indication of item redundancy or inconsistent item functioning. Moreover, temporal stability could be examined only in Study 1, as it included three waves of data collection. The results demonstrated good test–retest reliability of the Slovak version of the UWES over approximately four-month intervals. Our findings are consistent with those reported by Seppälä et al. (2009), who found high stability over a substantially longer time span of three years. At the same time, this level of stability raises an important conceptual question: to what extent does the observed stability in work engagement reflect enduring individual characteristics rather than stable features of the work context? Work engagement is defined as a relatively persistent yet dynamic state rather than a stable personality trait (Schaufeli & Salanova, 2007; Schaufeli et al., 2002). The observed stability may therefore partly reflect personal dispositions, but it may also be attributable to the relative stability of participants' work environments. According to the Job Demands–Resources model, engagement and burnout are shaped not only by personal resources but also by relatively stable job characteristics such as organizational culture, leadership, and workload (Bakker et al., 2023). Future research could disentangle these influences by examining individuals undergoing substantial job transitions and by controlling for stable personal resources in longitudinal designs.

In addition to these analyses, we evaluated the ultra-short UWES-3 to assess its suitability as a brief engagement measure. The scale also demonstrated strong psychometric performance, with high alpha and omega values and a MIC comparable to that of the full scale. This supports its potential usefulness in contexts where brief assessment tools are necessary. Our findings, therefore, align with Schaufeli et al. (2019), who recommend the UWES-3 primarily for large-scale surveys or settings where questionnaire length is a major constraint (e.g., national or international workforce surveys). Consistent with their conclusions, the brief scale performed well in our Slovak sample, but the full UWES-9 remains preferable for research and applied settings where a comprehensive assessment of work engagement is desired.

Lastly, we focused on validity analyses. In Study 1, convergent validity was supported through consistent negative associations between all work engagement measures (UWES-9, UWES-3, and UWES-9 subscales) and indicators of burnout, consistent with engagement being defined as the opposite side of burnout (Schaufeli et al., 2002; Schaufeli et al., 2006), and prior studies demonstrating a negative relationship between engagement and burnout (e.g., Schaufeli et al., 2002; Schaufeli et al., 2020; Sinval et al., 2022). Work engagement was also positively associated with self-reported work performance. The UWES-9 subscales of Vigor and Dedication showed the strongest negative correlations with overall burnout and specific symptoms, particularly Mental Distance. In contrast, Absorption showed weaker, though still significant, correlations with burnout dimensions.

The weakest correlations across measures were observed with Emotional Impairment and the UWES-9 total score. Similarly, work performance demonstrated the weakest (though still moderate and significant) association with the UWES-9 total score, while a stronger association was found with Vigor, Dedication, Absorption, and the UWES-3. Overall, the results support the convergent validity of both forms of the UWES, with evidence favoring the multidimensional structure, as indicated by factorial analyses from both studies and previous research (Schaufeli et al., 2006).

Study 2 further supported convergent validity through significant positive correlations between all UWES measures and the ISA engagement scale and its subdimensions. A consistent pattern emerged across measures: the strongest associations were observed with

Affective Engagement, followed by the overall ISA score, then Intellectual Engagement, and finally Social Engagement. Both the UWES-9 and UWES-3 displayed similar effect sizes and patterns of associations with the ISA dimensions, as did the Vigor and Dedication subscales. Absorption showed the weakest correlations with ISA dimensions, particularly with Social Engagement. These patterns are in line with the results from Tauetsile (2019) and Stoeber et al. (2013), although their comparisons focused on the UWES-9 and its subscales and not the UWES-3.

Limitations and Future Research

Several limitations should be considered when interpreting the findings of this research. Firstly, we allowed two residual correlations to improve model fit in Study 2. Even though it is not generally recommended, we consider it defensible, as it was grounded in theory (Schaufeli et al., 2019) and supported by results from Study 1.

Second, although we explored a bifactor structure in Study 1, it produced a Heywood case and was therefore not retained. As a result, conclusions regarding the presence and strength of a general engagement factor are limited.

Third, while most model fit indices supported the three-factor solution, the chi-square statistic was significant in all models, and the RMSEA for the unidimensional structure in Study 1 exceeded recommended thresholds.

Fourth, the study relied exclusively on self-report measures, including a single-item indicator of work performance. Although self-reports are common in organizational research and efficient for large samples, they are susceptible to common-method bias. Future research would benefit from incorporating other theoretically relevant constructs (e.g., job demands, job resources, burnout, or turnover intentions) and more objective physiological or behavioral indicators (e.g., biomarkers of stress, absenteeism records).

Finally, because data were collected through the same agency using anonymous participation, it was not possible to verify whether some individuals participated in both studies. Although the studies were treated as independent, potential overlap cannot be fully excluded and may have contributed to the replication of the results.

Conclusion

In summary, findings from both studies provide strong support for the use of the UWES in the Slovak context. The three-factor structure consistently demonstrated superior model fit compared to a unidimensional model, and discriminant validity analyses further supported the distinction between Vigor, Dedication, and Absorption, indicating that the multidimensional conceptualization of engagement is preferable. Measurement invariance analyses showed that the UWES-9 operates equivalently across men and women in both studies and across occupational levels in Study 2, allowing for score comparisons in future studies. Reliability estimates were high across the UWES-9 and its subscales, and the UWES-3, and convergent validity was supported by meaningful associations with burnout, work performance, and another engagement measure. Overall, the results suggest that the UWES-9 is a reliable and valid tool for assessing work engagement in Slovakia, with the UWES-3 serving as a practical alternative when using a longer scale is not feasible.

Acknowledgement

This research was supported by the Slovak Research and Development Agency under Grant APVV-23-0548. The authors declare no known

conflicts of interest. All study materials, including anonymized raw data, codebooks, R analysis scripts, and supplementary tables for both Study 1 and Study 2, are openly available on the Open Science Framework (OSF): https://osf.io/surpg

Authors' ORCID
Ester Nosáľová
https://orcid.org/0009-0008-6238-6953
Simona Dudová
https://orcid.org/0000-0002-8826-4128

References

Bakker, A. B., Albrecht, S. L., & Leiter, M. P. (2011). Key questions regarding work engagement. *European Journal of Work and Organizational Psychology*, 20(1), 4–28. https://doi.org/10.1080/1359432X.2010.485352

Bakker, A. B., Demerouti, E., & Sanz-Vergel, A. (2023). Job demands–resources theory: Ten years later. *Annual Review of Organizational Psychology and Organizational Behavior*, 10, 25–53. https://doi.org/10.1146/annurev-orgpsych-120920-053933

Balducci, C., Fraccaroli, F., & Schaufeli, W. B. (2010). Psychometric properties of the Italian version of the Utrecht Work Engagement Scale (UWES-9). *European Journal of Psychological Assessment, 26*(2). https://doi.org/10.1027/1015-5759/a000020

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). Sage.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261–304.

Carmona-Halty, M. A., Schaufeli, W. B., & Salanova, M. (2019). The Utrecht Work Engagement Scale for students (UWES–9s): Factorial validity, reliability, and measurement invariance in a Chilean sample of undergraduate university students. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01017

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Christian, M. S., & Slaughter, J. E. (2007). *Work engagement: A meta-analytic review and directions for research in an emerging area*. Paper presented at the Sixty-Seventh Annual Meeting of the Academy of Management, Philadelphia, PA. https://doi.org/10.5465/ambpp.2007.26536346

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience* (p. 1). New York: Harper & Row.

Demerouti, E., & Fullagar, C. J. (2016). Experiencing flow in the workplace and what individuals and organizations can do to foster it. In *The Fulfilling Workplace* (pp. 71–90). Routledge.

Dimitriadou, S., Lavidas, K., Karalis, T., & Ravanis, K. (2020). Study engagement in University Students: A confirmatory factor analysis of the Utrecht Work Engagement Scale with Greek students. *Journal of Well-Being Assessment*, *4*(3), 291–307. https://doi.org/10.1007/s41543-021-00035-7

Domínguez-Salas, S., Rodríguez-Domínguez, C., Arcos-Romero, A. I., Allande-Cussó, R., García-Iglesias, J. J., & Gómez-Salgado, J. (2022). Psychometric properties of the Utrecht Work Engagement Scale (UWES-9) in a sample of active health care professionals in Spain. *Psychology Research and Behavior Management*, 3461–3472. https://doi.org/10.2147/PRBM.S387242

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

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes. *Structural Equation Modeling, 11*(3), 320–341.

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Organizational Behavior*, *2*(2), 99–113.

Nerstad, C. G., Richardsen, A. M., & Martinussen, M. (2010). Factorial validity of the Utrecht Work Engagement Scale (UWES) across occupational groups in Norway. *Scandinavian Jour-*

nal of Psychology, *51*(4), 326–333. https://doi.org/10.1111/j.1467-9450.2009.00770.x

Rosseel, Y. (2012). lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

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*(4), 701–716. https://doi.org/10.1177/0013164405282471

Schaufeli, W. B., Martínez, I. M., Pinto, A. M., Salanova, M., & Bakker, A. B. (2002). Burnout and engagement in university students: A cross-national study. *Journal of Cross-Cultural Psychology, 33*(5), 464–481. https://doi.org/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., Shimazu, A., Hakanen, J., Salanova, M., & De Witte, H. (2019). An ultra-short measure for work engagement. *European Journal of Psychological Assessment*, *35*(4), 577–591. https://doi.org/10.1027/1015-5759/a000430

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., De Witte, H., & Desart, S. (2020). Manual Burnout Assessment Tool (BAT) – Version 2.0. KU Leuven, Belgium: Unpublished internal report.

Seppälä, P., Mauno, S., Feldt, T., Hakanen, J., Kinnunen, U., Tolvanen, A., & Schaufeli, W. (2009). The construct validity of the Utrecht Work Engagement Scale: Multisample and longitudinal evidence. *Journal of Happiness Studies*, *10*(4), 459–481. https://doi.org/10.1007/s10902-008-9100-y

Sinval, J., Pasian, S., Queirós, C., & Marôco, J. (2018). Brazil-Portugal transcultural adaptation of the UWES-9: Internal consistency, dimensionality, and measurement invariance. *Frontiers in Psychology*, *9*, 353. https://doi.org/10.3389/fpsyg.2018.00353

Sinval, J., Vazquez, A. C., Hutz, C. S., Schaufeli, W. B., & Silva, S. (2022). Burnout Assessment Tool (BAT): Validity evidence from Brazil and Portugal. *International Journal of Environmental Research and Public Health*, *19*(3), 1344. https://doi.org/10.3390/ijerph19031344

Soane, E., Truss, C., Alfes, K., Shantz, A., Rees, C., & Gatenby, M. (2012). Development and application of a new measure of employee engagement: The ISA Engagement Scale. *Human Resource Development International*, *15*(5), 529–547. https://doi.org/10.1080/13678868.2012.726542

Stoeber, J., Townley, J., & Davis, C. R. (2013). Comparing two work-engagement scales: Relationships with job satisfaction, organizational commitment, and workaholism (Research report, April 7, 2013). School of Psychology, University of Kent.

Tauetsile, J. (2019). Measuring employee engagement: Utrecht Work Engagement Scale (UWES) or Intellectual, Social and Affective Scale (ISA)? *Botswana Journal of Business*, *12*(1), 22–38.

Zecca, G., Györkös, C., Becker, J., Massoudi, K., de Bruin, G. P., & Rossier, J. (2015). Validation of the French Utrecht Work Engagement Scale and its relationship with personality traits and impulsivity. *European Review of Applied Psychology*, *65*(1), 19–28. https://doi.org/10.1016/j.erap.2014.10.003

Appendix

Slovak Translation of the Utrecht Work Engagement Scale

Inštrukcia

Nasledujúce výroky sa týkajú toho, ako sa cítite v práci. Pozorne si prečítajte každý výrok a rozhodnite sa, či sa takto niekedy cítite vo vašej práci. Ak ste tento pocit nikdy nemali, uveďte na stupnici číslo „0“ (nula). Ak ste tento pocit niekedy mali, označte výrok číslom od 1 do 6, podľa toho, ktoré číslo, najlepšie vystihuje, ako často sa váš pocit opakuje.

Položky

1. Vo svojej práci prekypujem energiou *(Vitalita)* *
2. Vo svojej práci sa cítim statný/á a plný/á energie *(Vitalita)*
3. Som nadšený/á zo svojej práce *(Oddanosť)* *
4. Moja práca ma inšpiruje *(Oddanosť)*
5. Keď ráno vstávam, teším sa na prácu *(Vitalita)*
6. Cítim sa šťastný/á, keď usilovne pracujem *(Pohltenie)*
7. Som hrdý/á na prácu, ktorú vykonávam *(Oddanosť)*
8. Som plne pohrúžený/á do svojej práce *(Pohltenie)* *
9. Moja práca ma úplne pohlcuje *(Pohltenie)*

Odpoveďová škála

0 – Nikdy
1 – Takmer nikdy (Niekoľkokrát do roka alebo menej)
2 – Zriedkavo (Raz za mesiac alebo menej)
3 – Niekedy (Niekoľkokrát za mesiac)
4 – Často (Raz za týždeň)
5 – Veľmi často (Niekoľkokrát za týždeň)
6 – Vždy (Každý deň)

Poznámka

* = položky UWES-3. Pôvodná škála (Schaufeli et al., 2006) bola preložená spätným prekladom. Je možné rátať celkové skóre, ktoré je súčtom skóre zo všetkých položiek. Toto celkové skóre poskytuje všeobecný prehľad o úrovni angažovanosti jednotlivca v práci. Okrem toho sa môžu analyzovať aj skóre pre jednotlivé dimenzie (vitalita, oddanosť a pohltenie), pre detailnejší pohľad na konkrétne aspekty angažovanosti.