

Validating the Indonesian Version of UWES-17: A Rasch Model Analysis of Work Engagement

Abdul Rahman Shaleh¹, Bahrul Hayat¹, Arung Samundra¹

Syarif Hidayatullah State Islamic University Jakarta¹

abd.rahman.shaleh@uinjkt.ac.id

Abstract

Work engagement is a positive psychology construct widely sought after and developed in various countries. Psychometric validation requires statistical analysis to measure and differentiate individuals across different cultures. This study aims to validate the Utrecht Work Engagement Scale-17 (UWES-17) instrument on hospitality workers in state-owned companies in Indonesia. The study population comprised 3185 individuals, representing various positions and types of hospitality service jobs within the land transportation sector of a state-owned enterprise. A non-probability sampling approach, specifically convenience sampling, was employed, selecting participants based on their accessibility and willingness to participate. The participants, aged between 18 and 62, came from diverse fields and work areas. Analysis of item responses in the test was the Rasch Model with a study focus on item fit and person fit from the UWES-17 Scale. In addition, assumption tests of unidimensionality and local independence were performed to ensure the validity of the measurement model. The scale's reliability was also assessed to evaluate the instrument's consistency. The results showed that the items were overall fit, but the criteria for person fit were unmet. This is due to the tendency for uniformity and too many diverse positions and cultural tendencies of state-owned companies. The implication is that further item development and confirmation are needed on the constituent aspects, especially the absorption constituent items. Additionally, developing a more consistent shorter version or discovering other dimensions of work engagement that are specific to collectivist societies may be possible.

Keywords: Rasch Model, Work Engagement, UWES-17's validation, Indonesian Context

Abstrak

Totalitas kerja merupakan konstruk psikologi positif yang banyak dicari dan dikembangkan di berbagai negara. Validasi psikometrik memerlukan analisis statistik untuk mengukur dan membedakan individu di berbagai budaya. Penelitian ini bertujuan untuk memvalidasi instrumen Utrecht Work Engagement Scale-17 (UWES-17) pada pekerja perhotelan di perusahaan milik negara di Indonesia. Populasi penelitian terdiri dari 3185 individu, yang mewakili berbagai posisi dan jenis pekerjaan layanan perhotelan dalam sektor transportasi darat di perusahaan milik negara. Sampel dipilih secara non-probability dengan menggunakan teknik pengambilan sampel convenience, berdasarkan aksesibilitas dan kemauan peserta untuk berpartisipasi selama periode pengumpulan data. Peserta, berusia antara 18 dan 62 tahun, berasal dari berbagai bidang dan area kerja. Analisis respons item dalam tes tersebut adalah Model Rasch dengan fokus penelitian pada kecocokan item dan kecocokan orang dari Skala UWES-17. Selain itu, uji asumsi unidimensionalitas dan independensi lokal dilakukan untuk memastikan validitas model pengukuran. Keandalan skala juga dinilai untuk mengevaluasi konsistensi instrumen. Hasil penelitian menunjukkan bahwa item-item tersebut overall fit, namun kriteria person fit belum terpenuhi. Hal ini disebabkan oleh adanya kecenderungan keseragaman dan terlalu banyaknya posisi dan kecenderungan budaya yang beragam pada perusahaan-perusahaan milik negara. Implikasinya adalah diperlukan pengembangan dan konfirmasi item lebih lanjut pada aspek-aspek penyusunnya, terutama item-item penyusun penyerapan. Selain itu, mengembangkan versi pendek yang lebih ajeg atau menemukan dimensi lain dari work engagement yang khas pada masyarakat kolektif dapat dilakukan.

Kata Kunci: Model Rasch, Totalitas kerja, Validasi UWES-17, Konteks Indonesia

Introduction

Positive organizational behavior is an approach within organizational psychology that emphasizes the development of positive psychological capacities such as self-efficacy, optimism, hope, and resilience. This concept highlights that these traits can be measured, developed, and managed to enhance work performance (Luthans, 2002; Luthans & Avolio, 2009). POB is closely associated with work engagement, a positive psychological state characterized by vigor, dedication, and full immersion in one's work. The work engagement construct, popularized by Schaufeli et al., refers to a positive, fulfilling state of mind related to work, characterized by vigor, dedication, and absorption. The term refers to an enthusiastic attitude towards achieving organizational goals, including all elements of attitudes and behaviors such as commitment, effort, enthusiasm, perseverance, and high motivation at work (Macey & Schneider, 2008). By using an explanatory model, POB raises positive attributions in work behavior (Bakker & Schaufeli, 2008; Cameron, 2003; Cameron et al., 2003) as something measurable, that can be developed, and used effectively to improve performance (Luthans, 2002). This concept has gained substantial global attention and is considered essential in assessing employee well-being, performance, and organizational commitment (Linley et al., 2006; Snyder & Lopez, 2023).

Among researchers and practitioners who use it, there are various understandings of work engagement. Industry and organizational practitioners see it as a must-have in the workplace and organizations, following academic inconsistencies in giving meaning to the construct (Macey & Schneider, 2008). For organizational practice, it is important to recognize that employees experience fluctuating engagement as they perform their jobs. One trend in the literature is that human resource scholars have begun to study the top-down impact of human resource management (HRM) systems and practices on employee work engagement. (Albrecht et al., 2015). For example, draw on several theoretical frameworks to propose an integrated strategic engagement model that considers organizational context factors, job context factors, and individual psychological and motivational factors that influence engagement. Saks et al. (2022) also suggest that engagement researchers may benefit from the ability-motivation-opportunity model to understand how HRM practices may influence engagement. Overall, there is a clear trend toward recognizing that HR practitioners must foster engagement beyond routine administration in many aspects of their managerial endeavors (Albrecht et al., 2015). This is because being fully engaged is a choice. It has been described through three work engagement elements: beliefs, incentives, and actions. Beliefs are related to trust and mental acceptance of the validity of something. This indicates that people are unlikely to engage if they do not trust their organization. Meanwhile, incentives are the expectation of rewards. This directs motivated action and effort. Actions will occur when he processes them mentally. Fully engaged workers can and want to act, believing their actions will create value that drives achieving the organization's vision.

However, the definition shows that workers who are deeply involved in their work will focus on achieving organizational goals and their role. Engagement here includes rational thoughts, emotions, and behaviors committed to devoting their abilities to achieving organizational goals. In addition, he is proud of his company and what he is doing in his work (Mundy, 2003). In short, work engagement is an outcome that is produced when values, passions, and actions are in line with the vision and values of the organization (Mundy, 2003; Saks, 2006). Its implementation in the workplace has resulted in a shift in organizational attention from talented workers to workers who can work totally. However, not everyone can be engaged in their organization and work. Some workers love their work so much, but do not love the organization where they work (Saks, 2006). As a result, when the energy, enthusiasm, and efficacy that are poured out when workers experience absorption are the positive side of the continuum of negative burnout experiences that are characterized by three interrelated dimensions: energy, involvement, and efficacy on one side as opposed to exhaustion, cynicism, and inefficacy on the other (Leiter & Maslach, 2003; Maslach, 2003, 2016).

The importance of this concept has attracted the interest of many researchers. Research on work engagement has been extensively studied in various contexts, among workers, in different countries and cultures. Several findings suggest a link to work engagement. For example, structural and psychological empowerment (Jafari et al., 2021), job satisfaction (Meena & Batra, 2022), subjective well-being (Farajat & Salah, 2023), organizations' health, well-being, and performance (Guest, 2014; W. B. Schaufeli & Salanova, 2010; Bakker, A. B., & Leiter, 2010; W. Schaufeli & Salanova, 2011). In the context of countries and cultures, it is found, for example, in Koreans (Choi et al., 2020), Filipinos (Abun et al., 2021), Japanese culture (Shimazu et al., 2008), Malaysians (Elghadi et al., 2024), Turkey (Yaz & Cemaloğlu, 2024), Portuguese (Oliveira et al., 2023), Spanish (Adanaqué-Bravo et al., 2023), Chinese and Pakistanis (Aman et al., 2023), and others. There are also studies based on various subjects, including hospitality workers (Anasori et al., 2023; Kanjanakan et al., 2023), teachers/educators (Gleeson et al., 2023; Van der Colff & Rothmann, 2009; Sasmoko et al., 2018), supply chain workers (Fahrizal et al., 2023), railway workers (Imanina et al., 2024; R. A. Rahman et al., 2023), millennial workers (Imanina et al., 2024), manufacturing workers (U. U. Rahman et al., 2017), leaders (Rothmann, 2017), public sector workers (Zahari & Kaliannan, 2023) and many others.

The exposure indicates that work engagement has been widely studied worldwide with the Utrecht Work Engagement Scale 17 (UWES-17). The scale comprises 17 items grouped into three subscales: vigor (6 items), dedication (5 items), and absorption (6 items). Each subscale reflects a different facet of engagement, forming a multidimensional view of how individuals experience their work. This scale has been used in various cultural contexts and has demonstrated good psychometric properties across different populations. Cultural nuances and socioeconomic factors can influence the applicability and validity of psychological instruments in different settings. This difference may be due to the influence of socio-cultural conditions in forming mental programs that become patterns for individuals, groups, and communities in how they think, feel, and act.

In recent years, Indonesia has also studied work engagement among its workforce. By conducting a thorough validity and reliability analysis, researchers can determine whether the scale is appropriate for assessing work engagement in Indonesia. There are several versions of UWES, namely UWES-17, UWES-9, and UWES-3. UWES-17 is a long and comprehensive version for measuring its three-dimensional components. UWES-9 is a short and efficient version while maintaining the three-dimensional structure and efficiency in measurement. UWES-3, an ultra-short version with only three items, is the most efficient measurement with a very similar relationship pattern to the overall indicators of the three dimensions. In Indonesia, adaptation and validation of the Utrecht Work Engagement Scale (UWES) have been carried out on several versions, 9 and 17 (Shaleh, 2016) adapted UWES-17 into Indonesian with the term "totalitas kerja" to avoid confusion with other terms such as work involvement or work engagement. Involving 586 samples of lecturers at nine state universities in Java, the CFA analysis found that the three-dimensional structure could be applied (Sidharta, 2018), using construction worker participants in Bandung City, also found that UWES-17 could be used well. (Kristiana et al., 2019) adapted the UWES-9 into Indonesian and tested it on 100 inclusive schoolteachers in Solo, Semarang, Surabaya, and Malang and, through CFA analysis, found that this version had good goodness of fit values. Meanwhile, information regarding the adaptation and validation of the UWES-3 in the Indonesian context is still limited. In contrast, in the Indonesian context, Helmi et al., (2020) found a fourth dimension of work engagement.

In this regard, although the Utrecht Work Engagement Scale (UWES) has been widely validated in Indonesia, recent studies have shown that revalidation is still needed to ensure the suitability and accuracy of the instrument in a particular cultural context. There are several reasons. Firstly, UWES-17 is preferred because of its more comprehensive nature in measuring work engagement. (TATHA et al., 2024) showed that although the UWES-9 is more efficient in practice, the UWES-17 is still recommended for use in studies that require comprehensive measurement of work engagement dimensions. Similarly, research by (Gwamanda et al., 2024) found that the factor structure of the

UWES may not be completely stable across all cultural contexts, so revalidation is needed to ensure the construct validity and reliability of the instrument across populations. Secondly, some studies still use CFA based on Classical Test Theory (CTT). However, the CTT approach has several limitations, including that both person and item parameters depend entirely on the sample group (Fan, 1998). CTT is sensitive to the sample used; when the sample changes, the estimates obtained from the measurement will also change (Courville, 2004). It is relatively easy for test data to meet the required assumptions (Erguven, 2013). Third, this study focuses on UWES-17. Although a Rasch analysis of the Indonesian version of the UWES-9 has previously been conducted by (Kristiana et al., 2019), the study was limited to the short version and a sample of inclusive schoolteachers. Likewise, (Sidharta, 2018) focused on workers in Bandung. Additionally, UWES-9 requires additional consideration in measuring other factors. Although this version is short, efficient, and easy to use, some studies suggest that the long version may be more appropriate in describing the overall dimensions of work engagement (Seppälä et al., 2009). The study used the UWES-17 with participants who were diverse in terms of work field, ethnicity, and work area.

This study uses Rasch Modeling in the analysis. Rasch modeling is part of modern test theory that reflects the “philosophy” of objective measurement, which goes far beyond just data analysis methods. Objective measurement refers to a type of measurement that is not dependent on who is being measured (test-dependent scoring), or in other words, a model of measurement that is not affected by sample-related issues (sample-bound). This means that the estimated ability level of an individual in Rasch modeling does not depend on the items presented. Rasch modeling analyzes how participants respond to each item with either a positive (high level) or negative (low level) response. Additionally, Rasch modeling uses item difficulty parameters to calculate the characteristics of the response. Therefore, latent scores may differ between two test-takers with the same raw score but different response patterns (Reeves & Marbach-Ad, 2016). Furthermore, Rasch modeling has several advantages that make it a sample-independent measurement theory, such as specific objectivity, parameter separation, and sufficiency (Mead, 2008). Specific objectivity ensures that the measurement of a person or object is not dependent on the items used during the measurement process. Parameter separation refers to separating person and item parameters in response behavior. This means that there is no direct relationship between item difficulty and the individual’s ability parameter in the estimation process. Sufficiency is the adequacy of the total responses in a measurement category to estimate the individual’s ability or trait in responding to that category.

Several assumptions in Rasch modeling must be met, namely: unidimensionality and local independence (Ayala, 2009). Unidimensionality is the most critical and fundamental assumption of the Rasch model (Chou & Wang, 2010). The unidimensionality assumption states that a measurement model should measure only a single construct. Testing this assumption is crucial to ensure that UWES-17 truly measures only work engagement. Next is the assumption of local independence, which means that responses to one item are not dependent on responses to other items (Ayala, 2009). Local independence implies no correlation between item measurement errors (Chou & Wang, 2010). This study uses Rasch modeling to verify that the items in UWES-17 meet these assumptions (unidimensionality and local independence) when adapting it to Bahasa Indonesia. Current psychometric practices encourage the use of Rasch modeling to provide additional comprehensive information in evaluating or representing item and sample characteristics within an instrument (Distefano et al., 2017). The explanation above specifically motivates the researcher to conduct an in-depth study on testing the assumptions of unidimensionality and local independence, evaluating item fit using Rasch modeling, testing person and item reliability, as well as examining potential item bias among Indonesian workers.

Methods

Participant

The study was quantitative, using a cross-sectional survey design and a psychometric approach to explore the internal psychometric properties of psychological scales. The population in this study comprises all employees registered with a state-owned industry (BUMN) operating in the land transportation sector. According to the company's official website, the total number of employees is 24,319. This study utilized a non-probability sampling approach, specifically employing convenience sampling. This method was selected due to its practicality and efficiency in accessing participants who were readily available and willing to participate, aligning with the study's time and resource constraints. Data was collected through an electronic form (Google Form) distributed to the company's employees. As a result, 3,185 respondents, consisting of various positions and types of hospitality jobs, agreed to participate in this study. The subjects consisted of 2782 men and 403 women. In terms of age, there were 260 people aged 18-20 years, 1824 (21-30 years), 813 (31-40 years), 242 (41-50 years), 42 (51-62 years), and one person did not report. Regarding worker status, there were 2968 contract employees, 212 permanent employees, and five daily casual workers. The subjects also consisted of various types of jobs, including 141 people (parking attendants), 451 (restaurant workers), 1440 (security/guards), 776 (cleaning service), and 377 office staff from various levels. In addition, participants came from the Jakarta head office area, all branches in Indonesia from Area 1 to Area 11.

Instrument

The UWES-17 instrument (Schaufeli et al., 2002) was first translated and content validated by language and psychology experts. The UWES-17 is a self-report questionnaire consisting of 17 items covering three subscales: passion (six items, for example, 'I am very passionate about my work'), dedication (five items, for example, 'My work inspires me'), and absorption (six items, for example, 'I feel happy when I am absorbed in my work'). All items are rated on a five-point frequency rating scale. The UWES-17 instrument (Schaufeli et al., 2002) was first translated and content-validated by language and psychology experts. The adaptation process followed established guidelines for cross-cultural instrument validation. Referring to the cross-cultural adaptation guidelines proposed by Beaton et al. (2000) and Sousa & Rojjanasrirat (2011), the process commenced with a forward translation from English to Bahasa Indonesia. Subsequently, a back-translation into English was performed to ensure the accuracy and equivalence of the translated content. The resulting translations were then meticulously reviewed by two bilingual experts with backgrounds in psychology and linguistics to ensure semantic and conceptual equivalence between the original and translated versions. This rigorous approach aims to maintain the validity and reliability of the instrument across different cultural contexts by achieving semantic, idiomatic, experiential, and conceptual equivalence.

Data Analysis

This study used the Rasch Model to analyze the psychometric properties of the UWES-17, focusing on items and person fit. Prior to this, Confirmatory Factor Analysis (CFA) was conducted to ensure that the measuring instrument used correctly measures what is intended to be measured, namely Work Engagement, using Mplus Version 7.11. In confirmatory factor analysis (CFA), the use of Maximum Likelihood (ML) assumes that the observed indicators follow a multivariate and continuous normal distribution, an assumption that is not appropriate for ordinal observed variables. In this case, the researcher used the MLMV Estimator, introducing modifications so that the statistical test has a sample distribution that is in line with the mean and variance of the chi-square distribution with relevant degrees of freedom (Asparouhov & Muthén, 2021). The research simulation showed that MLMV estimation is generally more effective than the MLM method based on the error ratio in Type I of the omnibus model. The MLMV test statistics perform well in most situations, except in large models involving extreme non-normality and small sample sizes (Maydeu-Olivares, 2017b). Furthermore, the Rating Scale Model (RSM) was used in this study. RSM asserts that a group of items have a similar

rating scale structure. This model is based on surveys or attitude questions where respondents are given the same response to several items, which is suitable for Likert-type scales, to estimate item difficulty and person ability on a logit scale. Preliminary tests assessed unidimensionality and local independence, ensuring the suitability of the data for Rasch modeling. Winsteps software facilitated the analysis, providing item fit statistics, person reliability, and separation indices. Further analysis included Differential Item Functioning (DIF) to examine potential biases across demographic groups. The Wright Map was employed to visually compare item difficulty and respondent ability levels, aiding in interpreting the instrument's effectiveness. Overall, this methodological approach ensured a comprehensive evaluation of the UWES-17's measurement properties, confirming its validity and reliability within the studied context.

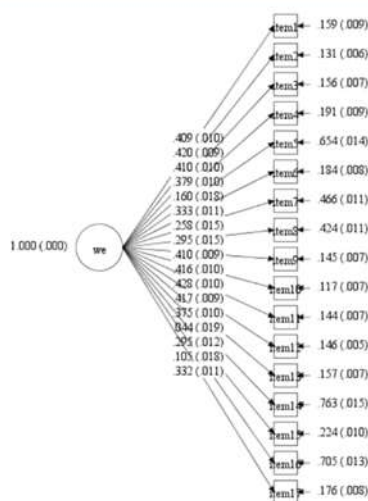
Results

Data Cleaning

The 3185 respondents obtained in this study were obtained through Google Forms. Before being analyzed, the researcher checked whether the data contained outliers by conducting a person fit analysis using the Rasch model. A total of 1547 respondents were discarded because they did not match the person fit order criteria, so this study only used 1638 respondents.

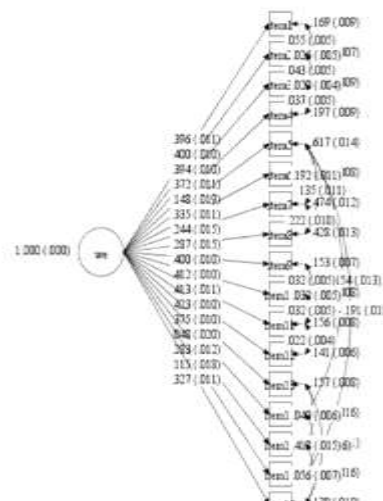
Confirmatory Factor Analysis

As an initial step in this study, a confirmatory factor analysis (CFA) was conducted. From the analysis conducted, the chi-square value (χ^2) = 2667.252, Comparative Fit Index (CFI) = 0.774, Tucker-Lewis Index (TLI) = 0.742, Standardized Root Mean Square Residual (SRMR) = 0.049 and Root Mean Square Error of Approximation (RMSEA) = 0.039 were obtained (Figure 1).



Sources: Personal data (2025).

Figure 1. Work Engagement's CFA Model



Sources: Personal data (2025).

Figure 2. Work Engagement's CFA Model

Because the model does not fit the data. The next step, the researcher modified the model, where the measurement errors in several items were freed from each other, the CFA results show that the proposed measurement model has a good fit with the empirical data. Model fit indices such as chi-square (χ^2) = 599.233, Comparative Fit Index (CFI) = 0.956, TLI = 0.94, Standardized Root Mean Square Residual (SRMR) = 0.049 and Root Mean Square Error of Approximation (RMSEA) = 0.039 indicate values that are in accordance with standards accepted in the literature (Figure 2). A CFI value greater than 0.90 and an RMSEA smaller than 0.05 (Umar & Nisa, 2020) indicate that the model has

an adequate level of fit. In addition, the factors tested in the measurement model show significant factor loadings, strengthening the construct validity of the measuring instrument used.

Overall, the results of this CFA provide support for the proposed theoretical structure and confirm the reliability and validity of the instrument in measuring the intended construct. These results are important to ensure that the data used in this study reflect concepts that are in accordance with existing theory, thus allowing for stronger generalization and interpretation in the context of further research. The next step is to see whether or not the item is significant in measuring what is to be measured, as well as determining whether or not certain items need to be dropped. In this case, what is tested is the null hypothesis about the factor loading coefficient of the item. The test is carried out by looking at the *t* value for each factor loading coefficient; if the *t* value > 1.96, it means that the item is significant, and vice versa. From Table 1, it can be seen that there are 17 items that have significant values (*t*-value > 1.96), and all coefficients are already positively loaded. This means that all factor loading coefficients of the items are valid.

Table 1. Factor Loadings of the Utrecht Work Engagement Scale-17 Items

Item	Estimate	S.E.	Est./S.E.	P-Value	Description
Item 1	0.694	0.016	42.860	0.000	V
Item 2	0.722	0.014	51.864	0.000	V
Item 3	0.693	0.015	44.910	0.000	V
Item 4	0.643	0.016	39.012	0.000	V
Item 5	0.185	0.023	8.060	0.000	V
Item 6	0.618	0.017	36.786	0.000	V
Item 7	0.334	0.020	16.508	0.000	V
Item 8	0.402	0.020	19.779	0.000	V
Item 9	0.714	0.014	52.014	0.000	V
Item 10	0.764	0.015	50.195	0.000	V
Item 11	0.723	0.016	46.551	0.000	V
Item 12	0.747	0.011	65.553	0.000	V
Item 13	0.688	0.015	44.926	0.000	V
Item 14	0.054	0.023	2.409	0.016	V
Item 15	0.507	0.021	23.985	0.000	V
Item 16	0.136	0.022	6.194	0.000	V
Item 17	0.611	0.019	31.651	0.000	V

Description: V sign = significant ($t > 1.96$); X = not significant

Sources: Personal Data (2025).

Unidimensionality

The first assumption that must be met in the Rasch model is unidimensional. In this case, to find out whether this measurement is said to be unidimensional by using the Principal Component Analysis of Residual (PCAR) method (Chou & Wang, 2010). PCAR functions to measure the extent to which the diversity of an instrument measures what is to be measured, which means that the purpose of unidimensional analysis is to see whether the items that have been created can measure a single dimension, namely Work Engagement. The model is said to have unidimensional as an objective measurement if the raw variance explained by the measure is >40% (Holster & Lake, 2016). Based on the test results above, the researcher obtained a raw variance explained by measures value of 49.1%

(Table 2), which means that the measuring instrument items in the measurement measure work engagement and have good unidimensional.

Table 2. Unidimensionality

Dimensi	Eigenvalues	Observed (%)	Expected
Total raw variance in the observation	33.4	100.0%	100%
Raw variance explained by measures	16.4	49.1%	50.7%
Raw variance explained by persons	13.7	41.0%	42.4%
Raw variance explained by items	2.7	8.1%	8.4%
Raw unexplained variance (total)	17.0	50.9%	100%
Unexplained variance 1st contrast	2.5	7.5%	14.7%
Unexplained variance 2nd contrast	1.9	5.7%	11.2%
Unexplained variance 3rd contrast	1.7	5.1%	9.9%
Unexplained variance 4th contrast	1.4	4.3%	8.5%
Unexplained variance in 5th constrast	1.1	3.3%	6.6%

Sources: Personal Data (2025).

Local Independence

The second assumption that must be met is local independence. This assumption requires that each set of items developed is not significantly related to other items or indicates that a person's response to an item is not influenced by the person's response to other items by looking at the residual correlation. In Rasch analysis, residual correlations measure the relationship between items after controlling for the respondents' overall ability. High residual correlation values can indicate a violation of the local independence assumption, which is crucial for the validity of the Rasch model. In this study, a residual correlation of 0.52 was found between item 14 ("I get carried away when I am working") and item 16 ("It is difficult to detach myself from my job"), indicating a significant potential for local item dependence (LID).

Table 3. Local Independence

Residual Correlin	Entry Number	Wording	Entry Number	Wording
0.52	14	Saya bahagia selama bekerja secara intens	16	Ketika sedang bekerja, saya larut dalam suasana

Sources: Personal Data (2025).

According to the literature, residual correlations exceeding 0.30 compared to the average residual correlation can indicate LID. This is supported by K. B. Christensen et al. (2017), Andrich (2018), and Yen (1984), who state that residual correlation values more than 0.30 above the average can signal local item dependence. LID occurs when responses to certain items are influenced by responses to other items, even after accounting for the underlying latent trait being measured. This interdependence violates the assumption of local independence inherent in the Rasch model, potentially leading to biased parameter estimates and undermining the unidimensionality of the scale. Consequently, LID can inflate reliability coefficients and distort the measurement of the intended construct. The presence of local item dependence (LID) can significantly compromise the validity and reliability of measurement instruments. These findings underscore the critical importance of assessing and addressing local item dependence to ensure the accuracy and validity of instruments analyzed using the Rasch model.

Item Difficulty

Table 4 contains the psychometric characteristics of work engagement resulting from the Rasch Model analysis. The table displays the item difficulty level on the logit scale, fit statistics, and step parameters. In this table, two items, namely items 5 and 14, did not match the Rasch Model fit limit with an acceptable fit value (0.5 - 1.5). The analysis indicates that one item does not fit the Rasch model and requires improvement. This misfit can be identified through the Infit and Outfit Mean Square (MNSQ) statistics (Linacre, 2002). The items are items 5 and 14 with the following statements: item 5 (I am enthusiastic about my job) and item 14 (Saya bahagia selama bekerja secara intens). Furthermore, the item difficulty level is in the symmetrical range, with the lowest value of 24.72 and the highest 65.35. Item number 5 is the most challenging item chosen by respondents, with a logit value of 65.35. On the other hand, item 16 reads, "Ketika sedang bekerja, saya larut dalam suasana ," with a logit value of 24.72. Item 16 is the most often chosen item by respondents. Items with an Outfit Mean Square (MNSQ) value below 0.5 are considered less productive for measurement purposes, even though they do not directly compromise the instrument's validity. In the table, no items with values below 0.5 were found. However, the presence of such items can artificially inflate reliability estimates and reduce the overall efficiency of the instrument. Low Outfit MNSQ values indicate that the item responses are overly predictable or redundant, suggesting that the item may not contribute meaningful information to the measurement construct. The results of the Rasch analysis indicate that two items do not fit the Rasch model and need to be revised.

Table 4. Item Difficulty

Entry Number	Measure	Infit Mnsq	Outfit Mnsq	Ptmea Corr.	Exact Obs %	Match Exp %
5	65.35	1.55	1.82	0.43	52.3	70.9
6	65.20	0.90	0.62	0.68	91.9	91.3
15	63.71	0.96	0.73	0.66	91.9	91.7
4	60.15	0.87	0.73	0.71	90.6	89.3
7	59.68	0.96	1.01	0.56	83.0	74.0
9	57.65	0.76	0.54	0.75	92.5	90.4
11	54.34	0.74	0.54	0.77	91.8	89.2
1	54.00	0.79	0.64	0.75	92.0	89.7
2	51.72	0.74	0.58	0.76	93.0	90.2
3	49.84	0.78	0.60	0.76	91.5	89.8
17	48.76	0.77	0.30	0.71	94.6	93.4
10	47.49	0.70	0.34	0.78	93.1	91.1
8	40.32	0.92	0.92	0.60	81.7	78.6
13	39.26	0.74	0.40	0.75	93.0	91.7
12	35.56	0.71	0.53	0.77	91.0	87.3
14	32.26	1.39	1.56	0.58	45.8	65.2
16	24.72	1.30	1.42	0.59	52.3	68.0

Sources: Personal Data (2025).

Item and Separation Reliability

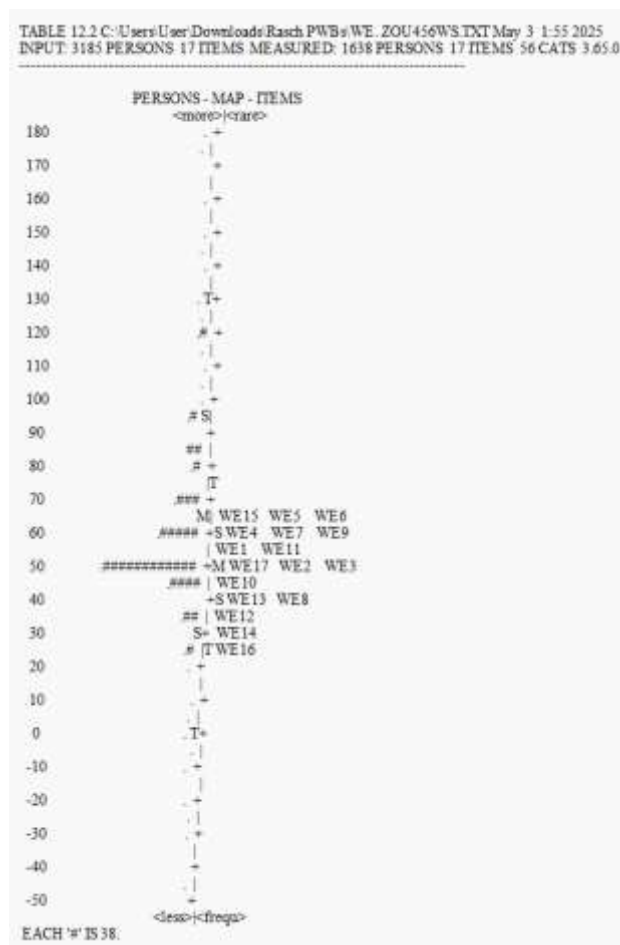
The concept of reliability differs between Classical Test Theory (CTT) and the Rasch model. In CTT, reliability is typically assessed as a single coefficient reflecting the consistency of test scores. In contrast, the Rasch model provides separate reliability estimates for persons and items, known as

Person Separation Reliability (PSR) and Item Separation Reliability (ISR), respectively (Wright & Masters, 1982). PSR evaluates how well the instrument distinguishes among individuals with varying levels of the latent trait, indicating the precision of person ability estimates. ISR assesses the extent to which the sample of respondents can reliably differentiate item difficulties, reflecting the stability of item calibrations. Furthermore, the Rasch model allows for the estimation of person abilities using the Weighted Likelihood Estimation (WLE) method. WLE provides less biased estimates of person parameters, especially for individuals with extreme scores, enhancing the accuracy of person reliability assessments (Adams, 2005). In summary, the Rasch model's approach to reliability offers a more nuanced understanding by separately evaluating the consistency of person and item measures, thereby providing deeper insights into the measurement properties of an instrument.

Person reliability can be assessed more accurately regarding scales in related constructs when person reliability in the Weighted Likelihood Estimate (WLE) is considered (Adams, 2005). In other words, the reliability of WLE describes some effective and reliable instruments for measuring the differences between work attitudes and well-being and the surrounding environment. Individual characteristics and qualifications are used to classify individuals. Low person reliability (less than 2) and person fitness (less than 0.8) with a relevant sample of individuals suggests that the tool may not be sensitive enough to distinguish between good and bad performers. Perhaps you need more items. Then, item reliability and separation are used to verify the item hierarchy. Low item separation (<3 = high, medium, low item difficulty, and item performance <0.9) indicates that the sample was not large enough to verify the instrument's item difficulty hierarchy (construct validity). The item reliability and separation results in this study were 1.00 and 17.50. In addition, the Rasch analysis yielded a person separation index of 1.98 and a person reliability coefficient of 0.80. These results indicate that the instrument can effectively distinguish between individuals across approximately three distinct levels of the latent trait being measured, demonstrating good measurement precision and consistency (Linacre, 2002). This shows that this instrument is not sensitive enough to distinguish between high and low performers and may require more items. The result of item separation reliability in this study was 1.00, which means that the sample of people is substantial to confirm the hierarchy of the instrument items' difficulty level. This shows that the reliability of the instruments of work engagement is very good.

Wright Maps and Person Measures

You can find additional information about item parameter estimation in Wright maps. Through the Wright Map in Figure 1, information on item difficulty level or previously presented in table form can be seen in graphical form. Wright MPAs provides information about the relationship between a person's latent trait level and the estimated item difficulty level on the same logit scale. In the picture below, the person's ability is higher than the question's difficulty level. This indicates that the targeting significantly differs, meaning the test is off target. In addition, the trickiest question chosen by respondents is WE5, and the most chosen by respondents is WE16.



Sources: Personal data (2025).

Figure 3. Wright Map

Differential Item Functioning

An item is said to have DIF if the item functions differently for respondents from different groups (e.g., education or gender), even though the respondents have the same abilities. DIF analysis is essential in the process. In order to assess development, it aims to ensure that the scores obtained from educational tests and psychological measurements cannot and do not reflect the same concept for all test participants or respondents (Walker, 2011).

Table 5. Differential Item Functioning

Item	Item-Trait Chi-Square Probability
1	0.6548
2	0.0608
3	0.4769
4	0.3618
5	0.0023
6	0.9158
7	0.0206
8	0.0002
9	0.0591
10	0.9343
11	0.5049
12	0.7548
13	0.1674
14	0.0712
15	0.3994
16	0.0007
17	0.7145

s*p<0.05

Sources: Personal Data (2025).

In DIF testing, researchers use gender as the grouping variable with the Chi-Square item-trait interaction method (Linacre, 2007). A significant bias indicating DIF is identified when the item probability value is less than 0.05. Based on the table below, items 5, 7, 8, and 16 show evidence of DIF.

Discussion

Based on the data analysis, this study included 1,547 out of 3,185 respondents after data cleaning for subsequent analysis using the Rasch model. The findings from the Rasch analysis revealed several key points: (1) the scale items effectively measure work engagement and exhibit unidimensionality; (2) items 14 and 16 indicate a significant potential for local item dependence (LID); (3) two items, namely items 5 and 14 did not fit the Rasch model and require improvement; (4) the item reliability and separation indices for the work engagement scale are very good; (5) the Wright map and person measure analysis indicate that respondents' abilities are higher than the difficulty levels of the items; and (6) four items, items 5, 7, 8, and 16, exhibit evidence of differential item functioning (DIF), suggesting potential bias across different respondent groups. These findings align with previous research indicating psychometric challenges in the UWES-17, such as item misfit and DIF, which cultural or linguistic factors may influence.

The confirmatory factor analysis (CFA) of the UWES-17 affirmed that the three dimensions—vigor, dedication, and absorption—significantly contribute to the construct of work engagement, aligning with Schaufeli and Bakker's (2004) conceptualization. However, the Rasch analysis revealed limitations in person reliability and person separation, indicating the instrument's reduced sensitivity in distinguishing varying levels of individual engagement. Additionally, violations of the local independence assumption

were identified, with high residual correlations among specific item pairs (e.g., items 14 & 16), suggesting potential redundancy and measurement bias.

These findings may be due to the abstract and multidimensional constructs. Non-cognitive scales often measure complex psychological constructs that are not directly observed. This can cause items in the scale to be more highly correlated, even after controlling for latent variables, thus violating the assumption of local independence (Yen, 1984, 1993). In addition, redundancy and similarity of items increase the risk of local item dependence (LID) because responses to one item can influence responses to other similar items. In addition, respondents may be influenced by previous questions. This is where cultural context and other demographic factors come into play (Wang & Willson, 2005; Yen, 1993). Based on these findings, several things can be done to develop the Indonesian version of the UWES, namely: testlets, item revisions, and alternative models. A testlet is a group of items with the same stimulus or context. Using testlets can help reduce LID by grouping interdependent items into one unit of analysis. This allows the model to handle local dependencies more effectively, as done by de Bruin et al. (2013). In addition, revision or deletion of problematic items can also be done. Items that show high LID or redundancy need to be revised or deleted to improve the quality of the measurement instrument. In this case, three things need to be done, namely: identifying items with high residual correlation values (eg, $Q_3 > 0.3$) that indicate local dependency, evaluating item content to detect similarity or redundancy, and revising or deleting problematic items to reduce dependency if revision is not possible (Lee et al., 2021; Shimazu et al., 2008). If the LID problem persists despite revision and use of testlet, then developing an alternative model can be a solution. This model development can be done by using a bifactor model to separate common variance from specific variance caused by LID, considering a multidimensional model if there is an indication that the construct being measured has more than one significant dimension as done by (Helmi et al., 2020), or exploring different factor structures to determine the model that best fits the data, for example developing a more robust condensed version as done by Kristiana et al., (2019), W. B. Schaufeli (2019); and W. B. Schaufeli et al. (2006).

This finding is likely because the differential Item Functioning (DIF) analysis is crucial in Rasch modeling to ensure that measurement instruments function equivalently across different groups, such as genders. DIF occurs when individuals from different groups, possessing the same underlying ability, have differing probabilities of responding correctly to a particular item. This can lead to biased estimates and affect the validity of the instrument. Differential Item Functioning (DIF) analysis within the Rasch model identifies items that function differently across groups—such as gender, age, education, ethnicity, language, and occupation—even when individuals possess the same underlying ability. This phenomenon can lead to biased estimates and affect the fairness of measurement instruments. The presence of DIF necessitates careful examination and, if necessary, revision of the affected items to ensure that the instrument measures the intended construct equitably across diverse groups. Addressing DIF is crucial for maintaining the validity and fairness of assessments, particularly in diverse populations (Linacre, 2011). In this study, the potential presence of DIF may be attributed to the diversity in gender, education, ethnicity, language, and occupational roles among respondents. Consequently, it is recommended to revise items 5, 7, 8, and 16 by rephrasing them into more universally comprehensible language to ensure clarity and fairness across all respondent groups. Specifically, items 5 and 14 are suggested for removal or replacement, as they may not effectively capture the intended construct across diverse populations. Items 5 and 14 can also be replaced with non-biased words or sentences.

In future studies, researchers can choose one of two things. First, using or adapting the short version such as the UWES-9 introduced by (W. B. Schaufeli et al., 2006) and Lee et al. (2021) in the Korean context. However, this requires considering several things. This is because, although the UWES-9 is efficient and easy to use, several studies have shown that this version has limitations in distinguishing in depth the three dimensions of work engagement—vigor, dedication, and absorption—and does not

include external factors such as job satisfaction, workload, and emotional exhaustion that can affect engagement, so it is recommended to use it together with other instruments to obtain a more comprehensive picture (Kulikowski, 2017; Seppälä et al., 2009; TATHA et al., 2024). In this context, for a deeper analysis the long version is still a better choice (TATHA et al., 2024).

Second, revising problematic items. Revising problematic items in the UWES scale is based on considerations to improve the construct validity, semantic clarity, and cultural appropriateness of the instrument. Items that exhibit negative or low factor loadings, have weak item-total correlations, or are ambiguous in meaning are often candidates for revision or deletion. Such revisions are important to ensure that each item accurately reflects the dimension being measured and is relevant in a specific cultural context, thereby improving the overall reliability and validity of the work engagement scale (de Bruin et al., 2013).

Furthermore, exploring the dimensionality of work engagement in different cultural contexts can provide valuable insights. For instance, a study by Helmi et al. (2020) in the Indonesian context identified a fourth dimension—contribution—through confirmatory factor analysis and the MIMIC model. This suggests that work engagement may encompass additional facets beyond the traditional three, highlighting the importance of cultural considerations in scale development and validation. Zahari & Kaliannan (2023) highlighted the influence of socially desirable responses in hierarchical public institutions like BUMN, where stability, collective values, and organizational conformity often shape employees' responses. These insights underscore the necessity for cultural adaptation and semantic refinement of the UWES-17 to enhance its validity and reliability within the Indonesian context. There is a pressing need to revise the items based on semantic and linguistic review, involving cross-cultural psychology and language experts. Items identified as misfitting should be re-evaluated qualitatively to ensure that statistical considerations and cultural and contextual understanding inform revisions. This is in line with the cross-cultural validation approach advocated by van de Vijver & Leung (2021). Therefore, future studies might focus on both refining the UWES for brevity and investigating potential additional dimensions to enhance its applicability and accuracy across diverse populations.

Conclusion

Based on the Rasch analysis conducted on 1,547 respondents, several psychometric issues with the UWES-17 were identified. While the scale demonstrated unidimensionality, items 14 and 16 showed significant local item dependence (LID), and items 5 and 14 did not fit the Rasch model, suggesting a need for revision or removal. Although item reliability and separation were high, lower person reliability and separation indicated limited sensitivity in distinguishing engagement levels. Differential Item Functioning (DIF) was found in items 5, 7, 8, and 16, implying potential bias across respondent subgroups. Differences in educational levels, occupational roles, and item wording (e.g., "absorption at work") may cause varying interpretations. Consistent with findings by Lee et al. (2021) and Helmi et al. (2020), these issues underscore the need for revisions. Recommended steps include modifying or removing problematic items, using testlets to reduce LID, conducting qualitative research to explore item interpretation, and applying cultural adaptations (van de Vijver & Leung, 2021) to enhance the scale's relevance and validity in diverse Indonesian contexts.

Acknowledgment

Thank you to the HRD Department of state-owned companies that provide land transportation services and have assisted in the data collection.

In compiling this manuscript, the author used limited ChatGPT assistance to improve sentence structure and readability of the text. All ideas, interpretations, and conclusions presented are entirely derived from the author's thoughts and analysis."

Conflict of Interest

The author declares that this research has no conflict of interest either in the preparation process or in writing this manuscript. The data found, processed, and analyzed, and the results presented are either internally or externally free from bias. This study is solely intended for academic and scientific purposes.

Authors Contribution

ARS conduct study design, write manuscripts, and participate in data analysis, BH contribute to advising and discussing the reporting and analysis of psychometric data, and AS conduct data analysis, looking for psychometric references, and discuss it.

References

- Abun, D., Menor, R. I., Catabagan, N. C., Magallanes, T., & Ranay, F. B. (2021). Organizational climate and work engagement of employees of divine word colleges in Ilocos Region, Philippines. *International Journal of Research in Business and Social Science* (2147- 4478), 10(1), 107–121. <https://doi.org/10.20525/ijrbs.v10i1.1017>
- Adams, E. S. (2005). Bayesian analysis of linear dominance hierarchies. *Animal Behaviour*, 69(5), 1191–1201. <https://doi.org/10.1016/j.anbehav.2004.08.011>
- Adanaqué-Bravo, I., Escobar-Segovia, K., Gómez-Salgado, J., García-Iglesias, J. J., Fagundo-Rivera, J., & Ruiz-Frutos, C. (2023). Relationship Between Psychological Distress, Burnout and Work Engagement in Workers During the COVID-19 Pandemic: A Systematic Review. *International Journal of Public Health*, 67(January), 1–11. <https://doi.org/10.3389/ijph.2022.1605605>
- Albrecht, S. L., Bakker, A. B., Gruman, J. A., Macey, W. H., & Saks, A. M. (2015). Employee engagement, human resource management practices and competitive advantage: An integrated approach. *Journal of Organizational Effectiveness*, 2(1), 7–35. <https://doi.org/10.1108/JOEPP-08-2014-0042>
- Aman, A., Rafiq, M., & Dastane, O. (2023). A cross-cultural comparison of work engagement in the relationships between trust climate – Job performance and turnover intention: Focusing China and Pakistan. *Heliyon*, 9(9), e19534. <https://doi.org/10.1016/j.heliyon.2023.e19534>
- Anasori, E., Soliman, M., & Costa, C. (2023). Workplace bullying, psychological distress, and work engagement in the hospitality industry: The moderating effect of self-compassion. *European Journal of Tourism Research*, 35, 1–18. <https://doi.org/10.54055/ejtr.v35i.3056>
- Andrich, D. (2018). Controlling response dependence in the measurement of change using the Rasch model. *Statistical Methods in Medical Research*, 27(12), 3709–3725. <https://doi.org/10.1177/0962280217710834>
- Asparouhov, T., & Muthén, B. (2021). *Bayesian analysis of latent variable models using Mplus*. 1–5. <https://www.statmodel.com/download/BayesAdvantages18.pdf>
- Ayala, R. J. de. (2009). Reviewed Work(s): Handbook of Statistics , Volume 26 : Psychometrics by C . R . Rao and Sandip Sinharay. *Journal of the American Statistical Association*, 104(487), <https://www.jstor.org/stable/40592308>
- Bakker, A. B., & Schaufeli, W. B. (2008). Positive organizational behavior: Engaged employees in flourishing organizations. *Journal of Organizational Behavior*, 29(2), 147–154.

<https://doi.org/10.1002/job.515>

- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*, 25(24), 3186–3191. <https://doi.org/10.1080/000163599428823>
- Cameron, K. S. (2003). Organizational virtuousness and performance. In *Positive organizational scholarship*. precisionmi.com. http://precisionmi.com/Materials/UniveralVirtuesMat/org_virtuousness_and_performance.pdf
- Cameron, K. S., Dutton, J. E., & ... (2003). An introduction to positive organizational scholarship. In *Positive arwana007.wordpress.com*. <https://arwana007.wordpress.com/wp-content/uploads/2013/11/positive-organizational-scholarship.pdf>
- Choi, M., Suh, C., Choi, S. P., Lee, C. K., & Son, B. C. (2020). Validation of the work engagement scale-3, used in the 5th Korean working conditions survey. *Annals of Occupational and Environmental Medicine*, 32(1), 1–10. <https://doi.org/10.35371/AOEM.2020.32.E27>
- Chou, Y. T., & Wang, W. C. (2010). Checking dimensionality in item response models with principal component analysis on standardized residuals. *Educational and Psychological Measurement*, 70(5), 717–731. <https://doi.org/10.1177/0013164410379322>
- Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied Psychological Measurement*, 41(3), 178–194. <https://doi.org/10.1177/0146621616677520>
- Courville, T. G. (2004). An empirical comparison of item response theory and classical test theory item/person statistics. *Dissertation, August*.
- de Bruin, G. P., Hill, C., Henn, C. M., & Muller, K. P. (2013). Dimensionality of the UWES-17: An item response modelling analysis. *SA Journal of Industrial Psychology*, 39(2 SPL), 1–8. <https://doi.org/10.4102/sajip.v39i2.1148>
- Distefano, C., Liu, J., & Greer, F. (2017). *Theory Methods : Examinations With a Teacher-Rating Scale*.
- Elghadi, A. M. O., Kassim, M. S., & Jusoh, A. (2024). Employee Engagement in the Malaysian Public Sector: the Moderating Effect of Job Demands on the Relationship Between Employee Participation, Employee Motivation, and Self-Efficacy. *International Journal of Business and Society*, 25(1), 180–200. <https://doi.org/10.33736/ijbs.6906.2024>
- Erguven, M. (2013). Two approaches in psychometric process: Classical test theory & item response theory. *Journal of Education*, 2(2), 23–30.
- Fahrizal, I., Santoso, B., & Budiono, A. (2023). Unlocking Work Engagement: How Leadership and Total Rewards Impact Employee Work Engagement Through the Mediating Role of Service Climate in Supply Chain and Logistic Company in Indonesia. *Jurnal Pamator : Jurnal Ilmiah Universitas Trunojoyo*, 16(2), 328–349. <https://doi.org/10.21107/pamator.v16i2.19791>
- Farajat, J., & Salah, A. A. (2023). Effect of positive organizational behavior on subjective well-being in the workplace in the tourism sector of Jordan. *Problems and Perspectives in Management*, 21(3), 736–752. [https://doi.org/10.21511/ppm.21\(3\).2023.57](https://doi.org/10.21511/ppm.21(3).2023.57)
- Gleeson, J., Cutler, B., Rickinson, M., Walsh, L., Ehrich, J., Cirkony, C., & Salisbury, M. (2023). School educators' engagement with research: an Australian Rasch validation study. *Educational Assessment, Evaluation and Accountability*, 35(2), 281–307. <https://doi.org/10.1007/s11092-023-09404-7>
- Guest, D. (2014). Employee engagement: a sceptical analysis. *Journal of Organizational Effectiveness*, 1(2), 141–156. <https://doi.org/10.1108/JOEPP-04-2014-0017>
- Gwamanda, N., Mahembe, B., & Jano, R. (2024). The psychometric properties of the Utrecht Work Engagement Scale on a Zimbabwean population. *SA Journal of Human Resource Management / SA*

Tydskrif Vir Menslikehulpbronbestuur, 22(0), a2492. <https://doi.org/10.1002/joe.22295>

- Helmi, A. F., Widhiarso, W., Marvianto, R. D., Priwati, A. R., Mustari, M. A., & Artikasari, Y. V. (2020). The Fourth Dimensions of the Job Demands-Resources (JD-R) Model of Work Engagement in Indonesian Context. *Jurnal Psikologi*, 47(3), 206. <https://doi.org/10.22146/jpsi.56682>
- Holster, T. A., & Lake, J. (2016). Guessing and the Rasch Model. *Language Assessment Quarterly*, 13(2), 124–141. <https://doi.org/10.1080/15434303.2016.1160096>
- Imanina, A. S., Rafiqana, N. M., & Hariani, M. (2024). *Analysing the influence of work-life balance on work with a qualitative approach PT Kereta Api Indonesia*. 2(2006), 55–58.
- Jafari, F., Salari, N., Hosseinian-Far, A., Abdi, A., & Ezatizadeh, N. (2021). Predicting positive organizational behavior based on structural and psychological empowerment among nurses. *Cost Effectiveness and Resource Allocation*, 19(1), 1–9. <https://doi.org/10.1186/s12962-021-00289-1>
- Kanjanakan, P., Zhu, D., Doan, T., & Kim, P. B. (2023). Taking Stock: A Meta-Analysis of Work Engagement in the Hospitality and Tourism Context. *Journal of Hospitality and Tourism Research*, 47(5), 851–876. <https://doi.org/10.1177/10963480211066958>
- Kristiana, I. F., Fajrianthi, F., & Purwono, U. (2019). Analisis Rasch Dalam Utrecht Work Engagement Scale-9 (Uwes-9) Versi Bahasa Indonesia. *Jurnal Psikologi*, 17(2), 204. <https://doi.org/10.14710/jp.17.2.204-217>
- Kulikowski, K. (2017). Do we all agree on how to measure work engagement? Factorial validity of Utrecht Work Engagement Scale as a standard measurement tool - A literature review. *International Journal of Occupational Medicine and Environmental Health*, 30(2), 161–175. <https://doi.org/10.13075/ijomeh.1896.00947>
- Lee, Y., Song, J. H., & Kim, S. J. (2021). Validation study of the Korean version of decent work scale. *European Journal of Training and Development*, 47(1), 43–57. <https://doi.org/10.1108/EJTD-03-2021-0040>
- Leiter, M. P., & Maslach, C. (2003). Areas of Worklife: a Structured Approach To Organizational Predictors of Job Burnout. *Research in Occupational Stress and Well Being*, 3(03), 91–134. [https://doi.org/10.1016/S1479-3555\(03\)03003-8](https://doi.org/10.1016/S1479-3555(03)03003-8)
- Linacre, J. M. (2002). Understanding Rasch measurement: Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement*, 3(1), 85–106. <https://pubmed.ncbi.nlm.nih.gov/11997586/>
- Linacre, J. M. (2011). *Winsteps Help for Rasch Analysis*. <http://homes.jcu.edu.au/~edtg/%5Cnpapers3:/publication/uuid/D56B724A-62FF-4D00-84E1-ECC888298B70>
- Linley, A. P., Joseph, S., Harrington, S., & Wood, A. M. (2006). Positive psychology: Past, present, and (possible) future. *Journal of Positive Psychology*, 1(1), 3–16. <https://doi.org/10.1080/17439760500372796>
- Luthans, F. (2002). The need for and meaning of positive organizational behavior. *Journal of Organizational Behavior*, 23(6), 695–706. <https://doi.org/10.1002/job.165>
- Luthans, F., & Avolio, B. J. (2009). The “point” of positive organizational behavior. *Journal of Organizational Behavior: The* <https://doi.org/10.1002/job.589>
- Macey, W. H., & Schneider, B. (2008). Engaged in Engagement: We Are Delighted We Did It. *Industrial and Organizational Psychology*, 1(1), 76–83. <https://doi.org/10.1111/j.1754-9434.2007.00016.x>
- Maslach, C. (2003). Job Burnout: New Directions in Research and Intervention. *Current Directions in Psychological Science*, 12(5), 189–192. <https://doi.org/10.1111/1467-8721.01258>

- Maslach, C. (2016). Understanding the burnout experience: Recent research and its implications for psychiatry. *World Psychiatry*, 15(2), 103–111. <https://doi.org/10.1002/wps.20311>
- Mead, R. (2008). A Rasch Primer: The Measurement Theory of Georg Rasch. *Psychometric Services Research Memorandum*, March. <http://www.edmeasurement.net/8226/Mead-2008-Rasch-primer.pdf>
- Meena, A. R., & Batra, G. (2022). Impact of Positive Organizational Behaviour on Job Satisfaction. *Editorial Board*, July. <https://fimt-ggsipu.org/pdf/Educator-Multidisciplinary-Journal.pdf#page=18>
- Mundy, M. E. (2003). *Faculty engagement in service-learning: individual & organizational factors at distinct institutional types*. Graduate School of Vanderbilt University.
- Oliveira, S., Carvalho, C., Pinto, A., de Moura, R. C., & Santos-Costa, P. (2023). Emotional labor, occupational identity and work engagement in Portuguese police officers. *International Journal of Human Resource Management*, 34(4), 768–804. <https://doi.org/10.1080/09585192.2022.2162345>
- Rahman, R. A., Pasaribu, F., Khair, H., & Tirtayasa, S. (2023). The Influence Of Talent Management And Knowledge Management On Employee Performance With Employee Engagement As The Intervening Variable On Pt Kereta Api Indonesia (Persero) Division Regional I Sumut. *Jurnal Ekonomi*, 12(04), 1–11. <http://ejournal.seaninstitute.or.id/index.php/Ekonomi>
- Rahman, U. U., Rehman, C. A., Imran, M. K., & Aslam, U. (2017). Does team orientation matter? Linking work engagement and relational psychological contract with performance. *Journal of Management Development*, 36(9), 1102–1113. <https://doi.org/10.1108/jmd-10-2016-0204>
- Reeves, T. D., & Marbach-Ad, G. (2016). Contemporary test validity in theory and practice: A primer for discipline-based education researchers. *CBE Life Sciences Education*, 15(1), 1–9. <https://doi.org/10.1187/cbe.15-08-0183>
- Rothmann, S. (2017). Employee Engagement. In *The Wiley Blackwell Handbook of the Psychology of Positivity and Strengths-Based Approaches at Work, First Edition*. Edited Employee Engagement. <https://doi.org/10.1002/9781118977620.ch1>
- Saks, A. M. (2006). Antecedents and consequences of employee engagement. *Journal of Managerial Psychology*, 21(7), 600–619. <https://doi.org/10.1108/02683940610690169>
- Saks, A. M., Gruman, J. A., & Zhang, Q. (2022). Organization engagement: a review and comparison to job engagement. *Journal of Organizational Effectiveness*, 9(1), 20–49. <https://doi.org/10.1108/JOEPP-12-2020-0253>
- Sasmoko, Abbas, B. S., Indrianti, Y., & Widhoyoko, S. A. (2018). Indonesian teacher engagement index: A rasch model analysis. *IOP Conference Series: Materials Science and Engineering*, 296(1), 1–8. <https://doi.org/10.1088/1757-899X/296/1/012027>
- Schaufeli, W. B. (2019). An ultra-short measure for work engagement: The UWES-3 validation across five countries. *European Journal of Psychological Assessment*, 35(4), 577–591. <https://doi.org/10.1027/1015-5759/a000430>
- 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., & Salanova, M. (2010). How to improve work engagement? In *Handbook of Employee Engagement: Perspectives, Issues, Research and Practice* (pp. 399–415). <https://doi.org/10.4337/9781849806374.00044>
- Schaufeli, W., & Salanova, M. (2011). Work engagement: On how to better catch a slippery concept. *European Journal of Work and Organizational Psychology*, 20(1), 39–46. <https://doi.org/10.1080/1359432X.2010.515981>
- 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>
- Shaleh, A. R. (2016). Analisis Faktor Skala Totalitas Kerja (Work Engagement). *SEMINAR ASEAN 2 Nd Psychology & Humanity*, 12–17.
- Shimazu, A., Schaufeli, W. B., Kosugi, S., Suzuki, A., Nashiwa, H., Kato, A., Sakamoto, M., Irimajiri, H., Amano, S., Hirohata, K., Goto, R., & Kitaoka-Higashiguchi, K. (2008). Work engagement in Japan: Validation of the Japanese version of the utrecht work engagement scale. *Applied Psychology*, 57(3), 510–523. <https://doi.org/10.1111/j.1464-0597.2008.00333.x>
- Sidharta, I. (2018). Validation on Utrecht Work Engagement Scale: Multisample and Multi Sectors Analysis. *Kontigensi : Jurnal Ilmiah Manajemen*, 6(2), 49–58. <https://doi.org/10.56457/jimk.v6i2.53>
- Snyder, C. R., & Lopez, S. J. (2023). The Future of Positive Psychology; A Declaration of Independence. *Handbook of Positive Psychology*, 751–767. <https://doi.org/10.1093/oso/9780195135336.003.0055>
- Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, 17(2), 268–274. <https://doi.org/10.1111/j.1365-2753.2010.01434.x>
- TATHA, O., SHIMAZU, A., WATANABE, K., KAWAKAMI, N., & SCHAUFELI, W. B. (2024). Measuring work engagement in Thailand: development and validation testing of the Utrecht Work Engagement Scale-Thai version (UWES-TH). *Industrial Health*, 62(3), 182–194. <https://doi.org/10.2486/indhealth.2023-0017>
- Umar, J., & Nisa, Y. F. (2020). Uji Validitas Konstruk dengan CFA dan Pelaporannya. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, 9(2), 1–11. <https://doi.org/10.15408/jp3i.v9i2.16964>
- van de Vijver, F. J. R., & Leung, K. (2021). Methods and Data Analysis for Cross-Cultural Research. *Methods and Data Analysis for Cross-Cultural Research*. <https://doi.org/10.1017/9781107415188>
- Van der Colff, J. J., & Rothmann, S. (2009). Occupational stress, sense of coherence, coping, burnout and work engagement of registered nurses in South Africa. *SA Journal of Industrial Psychology*, 35(1), 1–10. <https://doi.org/10.4102/sajip.v35i1.423>
- Walker, C. M. (2011). What's the DIF? why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment*, 29(4), 364–376. <https://doi.org/10.1177/0734282911406666>
- Wang, W. C., & Willson, M. (2005). Exploring local item dependence using a random-effects facet model. *Applied Psychological Measurement*, 29(4), 296–318. <https://doi.org/10.1177/0146621605276281>
- Wright, B. D., & Masters, G. N. (1982). *RATING SCALE ANALYSIS-n Item Response Modeling Approach*.
- Yaz, A. H., & Cemaloğlu, N. (2024). Tendencies of the doctoral theses on work engagement in Türkiye: a content analysis study. *Eximia Journal*, 13, 510–525. <https://doi.org/10.47577/eximia.v13i1.481>
- Yen, W. M. (1984). Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model. *Applied Psychological Measurement*, 8(2), 125–145. <https://doi.org/10.1177/014662168400800201>
- Yen, W. M. (1993). Scaling Performance Assessments: Strategies for Managing Local Item Dependence. *Journal of Educational Measurement*, 30(3), 187–213. <https://doi.org/10.1111/j.1745-3984.1993.tb00423.x>
- Zahari, N., & Kaliannan, M. (2023). Antecedents of Work Engagement in the Public Sector: A

Systematic Literature Review. *Review of Public Personnel Administration*, 43(3), 557–582.
<https://doi.org/10.1177/0734371X221106792>

Appendix

Table A. UWES-17 Instrument

No	UWES-1 in English	UWES-1 in Bahasa
1	At my work, I feel bursting with energy	Di tempat kerja, saya merasa penuh dengan energi
2	At my job, I feel strong and vigorous	Saya merasa sangat kuat dan bertenaga ketika mengerjakan pekerjaan saya
3	When I get up in the morning, I feel like going to work	Setiap pagi, saya berangkat bekerja dengan perasaan senang
4	I can continue working for very long periods at a time	Saya tahan berlama-lama dalam bekerja
5	At my job, I am very resilient, mentally	Saya tangguh dalam pekerjaan saya
6	At my work I always persevere, even when things do not go well*	Di tempat kerja, saya selalu tekun, bahkan saat sesuatu tidak berjalan dengan baik
7	I find the work that I do full of meaning and purpose	Menurut saya, pekerjaan saya penuh dengan makna dan tujuan
8	I am enthusiastic about my job	Saya antusias dengan pekerjaan saya
9	My job inspires me	Pekerjaan saya menginspirasi saya
10	I am proud on the work that I do	Saya bangga dengan pekerjaan yang saya lakukan
11	To me, my job is challenging	Pekerjaan saya merupakan hal yang membuat saya tertantang
12	Time flies when I'm working	Waktu berlalu dengan cepat ketika saya sedang bekerja
13	When I am working, I forget everything else around me	Saat bekerja, saya mengabaikan segala hal yang tidak terkait dengan pekerjaan
14	I feel happy when I am working intensely	Saya Bahagia selama bekerja secara intens
15	I am immersed in my work	Saya dapat menjadi lupa waktu ketika sedang bekerja
16	I get carried away when I'm working	Ketika sedang bekerja, saya larut dalam suasana
17	It is difficult to detach myself from my job	Sulit untuk mengalihkan perhatian saya saat sedang bekerja

Sources: Personal Data (2025).