Dimesionality of the Spanish short version of Utrecht Work Engagement Scale in Ecuador: Structural Considerations

Zuleima Santalla-Banderali1 and Jesús M. Alvarado2

**Abstract**

In this paper, we evaluate the factorial validity of the Spanish short version of the Utrecht Work Engagement Scale (UWES-9) in its adaptation to Ecuador and assess its predictive validity with respect to self-assessed work performance. A total of 229 employees from educational institutions in Ecuador participated. Using a model comparison analysis, the unidimensional model exhibited an excellent goodness of fit that was not improved by more complex models, and therefore is justified as an instrument of work engagement (WE). However, upon analyzing the correlation patterns of the overall score and the WE dimensions in relation to the task and contextual performance, including counterproductive behaviors, we conclude that, while the unidimensional model exhibits a good fit, the three-factor theoretical approach is significantly superior in that it maintains differential predictive validity for each theoretical dimension.

**Introduction**

In studying people's behavior in the workplace, scholars have shifted their attention to positive organizational behavior (Balducci et al., 2010) and the positive psychological states experienced by workers (Schaufeli, 2013). In this context, an aspect that has received considerable attention is work engagement. This is evidenced by the results obtained by Motyka (2018) in his systematic review of the literature, in which he analyzed 71 studies on engagement published in 25 countries between 2002 and 2017, observing that 70% of these studies had been published between 2014 and 2018.

However, the boom in research on work engagement is not evenly distributed across all countries and continents. In this connection, Motyka (2018) found that 20% of the studies included in his review had been conducted in the United States of America and the Netherlands, and 46% came from Europe; there were no studies conducted in Latin American countries.

The relevance of work engagement is evidenced in several theoretical models that include this construct as a mediator of the relationship between other variables and job performance. For example, Bakker (2011) proposed that work resources (e. g. social support from colleagues and supervisors, feedback on performance, the variety of skills required by the tasks to be performed and autonomy at work), and personal resources (e. g. optimism, self-efficacy, resilience and self-esteem) initiate a motivational process that results in an increase in work engagement, which in turn results in better performance. Similarly, Christian et al. (2011) posit in their conceptual model that engagement mediates the relationships between job characteristics, leadership, dispositional characteristics, and task and contextual performance.

Additionally, there is ample empirical evidence showing that there is a positive and statistically significant relationship between work engagement and job performance (Afacan-Findikli, 2015; Breevaart et al. , 2015; Christian et al, 2011; Dalal et al. , 2012; Gutermann et al. , 2017; Kim et al. , 2013; Lorente et al. , 2014; Lupano et al. , 2018; Motyka, 2018; Pongton & Suntrayuth, 2019; Qodariah et al. , 2019; Rich et al. , 2010). Thus, work engagement is considered a key antecedent of employee behavior (Bakker et al. , 2011; Bothma & Roodt, 2012), which contributes to the competitive advantage of organizations, increasing their financial performance and reputation (Karanges et al. , 2014; Shuck & Reio, 2011; Tampubolon, 2016).

Despite the above, work engagement is a complex construct and even today there is no consensus on its conceptualization (Chaudhary et al. , 2012; Wefald et al. , 2012), although several definitions that have been proposed could be considered complementary.

**Conceptualization of work engagement**

The complexity of work engagement (WE) is evident from the theoretical discussion surrounding it. Scholars have debated whether WE is a construct in itself or whether it overlaps with other constructs already used in the organizational field. In this sense, empirical evidence consistently shows that the WE is, in fact, it own construct distinct from others such as organisational commitment, job involvement, job satisfaction, organisational citizenship behaviour and extra-role behaviour (Bakker et al., 2011; Kataria et al., 2013; Saks, 2006). Similarly, the debate on whether or not WE is the polar opposite of burnout now also seems to have been resolved. Schaufeli (2013) and Schaufeli and Salanova (2011) explain that these two constructs are independent, have different structures and explain different aspects of the relationship that employees establish with their jobs (Schaufeli et al., 2002). Thus, although there is an inverse relationship between the two (Schaufeli, 2013), it is far from perfect, so that the fact that a person does not experience burnout does not mean that he or she feels engaged and vice versa (Schaufeli & Salanova, 2011).

The remaining theoretical discussion surrounds the uni- or multidimensionality of the construct. In this regard, Britt (1999, as cited in Welfald et al., 2012) and the Gallup Organization agree on understanding WE as a unidimensional construct. However, they disagree on how to define it. For Britt, the WE is a motivational state that refers to the extent to which the employee cares about their job performance, feeling responsible and committed to that performance (Welfald et al., 2012); while for the Gallup Organization the WE refers to the participation, satisfaction and enthusiasm of the individual for their work (Schaufeli, 2013; Shuck, 2011).

In contrast to the above, Kahn (1990), recognized as the first to apply the notion of engagement to the work context, presents it as a multidimensional construct consisting of three dimensions: (a) cognitive engagement; (b) emotional engagement; and (c) behavioral engagement. Moreover, Kahn defines WE as a psychological state in which employees "incorporate" their personal selves into their work performances, investing personal energy and experiencing an emotional connection to their work. This enables them to be connected to their job roles physically, cognitively, emotionally (Bakker, 2011; Schuck, 2011, 2013; Shuck & Reio, 2011). devote their personal resources to performing the tasks associated with the role they play within the organization (Christian et al. 2011), and focus their energy on organizational goals (Bakker, 2011).

For Kant, the cognitive engagement corresponds to the evaluation that employees make about whether the work they do is meaningful and safe, and whether they have adequate resources to do it (Shuck & Reio, 2011). The emotional engagement refers to the emotional bond that the person feels with their workplace and represents the individual's willingness to involve their personal resources in their work (Shuck & Reio, 2011). Finally, behavioral engagement is the intention to act for the sake of better performance, being the explicit manifestation of cognitive and emotional engagement (Shuck & Reio, 2011).

The view of WE as a multidimensional construct is shared by Schaufeli and colleagues, who define it as a positive, satisfying and active psychological state associated with the performance of daily work activities, characterized by vigor, dedication and absorption in their work (Bakker, 2011; Salanova-Soria & Schaufeli, 2004; Schaufeli, 2013, 2017; Schaufeli et al. , 2002), with vigor and dedication being the core dimensions (Schaufeli & Salanova, 2011). According to Schaufeli (2013, 2017), this notion is congruent with that of Kant (1990), in the sense that there is an equivalence between physical-energetic and vigor, emotional energy and dedication, and a cognitive component and absorption.

Vigor refers to the extent to which the worker experiences high levels of energy and physical and mental endurance while working (Bakker, 2011; Bakker et al., 2011, 2012; Schaufeli, 2013). That is, their desire to invest effort in the work they perform and to persist even in times of difficulty (Salanova-Soria & Schaufeli, 2004; Schaufeli, 2013, 2017; Schaufeli et al. , 2002).

Dedication refers to the degree to which the person feels involved and identified with the work they do and experiences a sense of importance, enthusiasm, inspiration, pride, and challenge (Bakker, 2011; Bakker et al. , 2011, 2012; Salanova-Soria & Schaufeli, 2004; Schaufeli, 2013, 2017; Schaufeli et al. , 2002).

Finally, absorption alludes to the extent to which the person is so focused and absorbed in their work (Bakker, 2011; Bakker et al. , 2011, 2012; Schaufeli, 2013), which facilitates the quick passing of time (Bakker, 2011; Schaufeli, 2013). High absorption entails a difficulty in distancing oneself from their work (Salanova-Soria & Schaufeli, 2004; Schaufeli, 2013, 2017; Schaufeli et al. , 2002; Simpson, 2009).

For Schaufeli and colleagues, the WE is relatively stable over time and is not focused on a particular event, object, person, or behavior (Salanova-Soria & Schaufeli, 2004; Schaufeli & Salanova, 2011; Schaufeli et al., 2002; Simpson, 2009). However, this does not mean that it cannot change depending on the characteristics of the job (Schaufeli & Salanova, 2011) and what happens throughout the day (Bakker, 2011). In fact, some researchers cited in Breevaart et al. (2012) have measured the construct on a daily basis, confirming that the WE reported by a particular person can fluctuate over time.

The iteration of WE proposed by Schaufeli and colleagues is the one mostly adopted by researchers, according to the results of the literature review conducted by Motyka (2018), and the one taken as a starting point in this study.

**Measurement of work engagement**

Although several instruments have been developed to measure WE, the most widely used measure (Rich et al., 2010; Schaufeli, 2013; Shuck, 2011; Welfald et al, 2012) is the Utrecht Work Engagement Scale (UWES) created by Schaufeli and colleagues based on their conceptualization of engagement as a multidimensional construct; an instrument which is even considered the standard measure of this construct (Kataria et al., 2013; Kulikowski, 2017; Lovakov et al., 2017; Rodríguez-Montalbán et al., 2014; Souza-Vázquez et al., 2017; Willmer et al., 2019). In fact, Motyka's (2018) systematic review found that in 82% of the publications analysed, the WE had been measured using all versions of the UWES.

Originally the UWES included 24 items (Schaufeli et al., 2002, 2006); but, after several psychometric evaluations and in order to achieve a more parsimonious measurement, the authors reduced its length to 17 items (Kataria et al., 2013; Rodríguez-Montalbán et al., 2014; Schaufeli et al., 2006): 6 measuring the vigor dimension; 5 for the dedication dimension; and another 6 for the absorption dimension. Nevertheless, we can find works with a version of the scale with 15 items, 5 for each of the dimensions, due to some results indicating inconsistency with two items in the UWES-17 across versions (Salanova et al., 2000).

Subsequently, Schaufeli et al. (2006) developed a nine-item version (UWES-9), three for each of the dimensions, which was validated in 10 countries with a sample of more than 14,000 people. This has become the most popular version, as according to Motyka (2018), it has been used used in 68% of the items analyzed in their literature review.

The UWES-9 has psychometric properties equivalent to its longer versions (Chaundhary et al., 2012; Kataria et al., 2013; Schaufeli et al., 2006), noting that it even has superior practical utility and construct validity than the UWES-17 (Schaufeli et al., 2006; Seppälä et al., 2009). In fact, in Kulikowski's (2017) review of 21 studies in which the construct validity of the UWES was assessed using confirmatory factor analysis (CFA), it was found that in eight of the nine studies in which the UWES-17 was compared to the UWES-9, the latter version showed better psychometric performance. In addition, scores on the UWES-9 are more stable over time and explain about 80% of the variance in scores on the UWES-17 (Kulikowski, 2017).

**Psychometric properties of the UWES-9**

The UWES-9 presents high internal consistency indices, reporting Crombach's Alpha coefficients for the full scale above 0.80 (Balducci et al., 2010; Chaundhary et al., 2012; Kataria et al., 2013; Klassen et al., 2012; Schaufeli et al., 2006; Villotti et al., 2014; Welfald et al., 2012; Willmer et al., 2019; Yusoff et al., 2013); and α values between 0.59 and 0.92 for the different dimensions (Balducci et al., 2010; Chaundhary et al., 2012; Hallberg & Schaufeli, 2006; Kataria et al., 2013; Klassen et al., 2012; Lovakov et al., 2017; Rodríguez-Montalbán et al., 2014; Schaufeli et al., 2006; Villotti et al., 2014; Welfald et al., 2012).

On the other hand, regarding the multidimensional structure of the UWES, there are wide discrepancies in the results obtained, which casts doubt on the generalizability of the notion of work engagement as a construct made up of the three dimensions proposed by Schaufeli and colleagues. In this sense, Schaufeli et al. (2006), Villotti et al. (2014) and Willmer et al. (2019) obtained correlations between the three dimensions greater than 0.70. Breevaart et al. (2012), De Bruin and Henn (2013), Hallberg and Schaufeli (2006), and Souza-Vázquez et al. (2017) found correlations greater than 0.80; and Balducci et al. (2010), Chaundhary et al. (2012), Gómez-Garbero et al. (2019), Lovakov et al. (2017), and Schaufeli et al. (2006) reported correlations greater than 0.90. Rodríguez-Montalbán et al. (2014) found that, on average, the three UWES subscales share 57% of the variance, again indicating that the theoretical dimensions of the UWES are highly correlated and suggesting unidimensionality.

Additionally, in Chile and using exploratory factor analysis (EFA), Müller et al. (2013) obtained two factors. In contrast, also using EFA, Chaundhary et al. (2012) in India, Souza-Vázquez et al. (2017) in Brazil, and Willmer et al. (2019) in Sweden found a one-dimensional solution; but, confirmatory factor analysis (CFA) showed that the three-factor model had slightly higher fit indices than the one-factor solution in the cases of India and Brazil.

The three-factor model has also shown acceptable fit rates in Uruguay (Gómez-Garbero et al., 2019), superior to the one-factor model in Italy (Balducci et al., 2010), in the Netherlands (Balducci et al., 2010; Breevaart et al., 2012) and with the multinational sample used by Schaufeli et al. (2006), and better than the one- and two-factor model (dedication+vigor, and absorption) in Pakistan (Yusoff et al., 2013). However, in studies such as Balducci et al. (2010) results indicative of a good fit of the three-factor model were obtained only when it was modified to allow for the error covariance between four pairs of items in the case of the Italian sample, and between two pairs of items in the Dutch sample. This correlation between errors suggests a possible content overlap between items (or redundancy problem).

In contrast to the above, the results of Souza-Vázquez et al. (2017) in Brazil, Hallberg and Schaufeli (2006) in Sweden, and Villotti et al. (2014) in Italy have shown non-statistically significant differences between the one-dimensional and three-factor models in terms of their fit indices. Additionally, Rodríguez-Montalbán et al. (2014) in Puerto Rico and Lovakov et al. (2017) in Russia found that the three-factor model presented a slightly better fit than the one-dimensional model, although the latter also had acceptable fit indices when the models were modified to allow for error covariance between three pairs of items (Lovakov et al., 2017), thus allowing for possible content redundancy.

Questions about the multidimensional structure of the UWES increase when considering that there are studies in which, unlike the previous ones, what has been observed is a superiority of the unidimensional model. For example, working with samples from Australia, Canada, China, Indonesia and Oman, Klassen et al. (2012) found that, considering together the data from the samples of the five countries and when the uni- and three-factor models were modified to allow for the error covariance between items, the fit indices of the one-factor model were slightly better than those of the three-factor model. Likewise, Fong and Ho (2015), using Bayesian structural equation modeling (BSEM) and specifying informative residual covariance priors, found that while all three models tested (one-factor, three-factor and partial bi-factor model) had adequate fit indices, the one-dimensional BSEM had a significantly lower BIC, indicating a better fit than the three-factor and partial bi-factor BSEM models. However, De Bruin and Henn (2013) found in South Africa that the model with the best fit indices was a partial bifactorial model which specifies a general factor which the authors named "work engagement" and which explained 89% of the variance, and two specific factors: vigor and absorption.

Finally, there are also results showing inadequate fit of the various models tested: one-factor, two-factor and three-factor (Kataria et al., 2013; Wefald et al., 2012; Willmer et al., 2019). In the case of Klassen et al. (2012), this was true when models were evaluated without allowing for correlation between errors. In the case of Fong and Ho (2015), none of the models tested (one-factor, three-factor and partial bifactor) presented adequate fit indices when the fit was assessed using the traditional ML approach of CFA. Finally, Hallberg and Schaufeli (2006) and Villotti et al. (2014) found that, for both the three-factor and unidimensional models, the RMSEA and χ2 fit indices indicated an unacceptable fit.

The discrepancy of results regarding the structure of the UWES has also been highlighted by Kataria et al. (2013) who found that the one-factor model had a better fit in 33.33% of the 10 studies reviewed; but, in another 33.33% a better fit of the three-factor solution was found, and in 20% neither solution was found to have adequate fit indices. More recently, Kulikowski (2017) reviewed 21 studies in which the factor structure of the original and the shortened version of the UWES, using CFA, was evaluated and confirmed the instability in the factor structure of the scale: in 23.81% of the studies the original three-factor structure was found, but in another 23.81% a unidimensional structure was found, and in 38.09% the results showed that both models (one-factor and three-factor) were equivalent.

This variability of results demonstrates that the UWES-9 may have a problem with the purported invariance of its factorial structure, demanding further examination (Kulikowski, 2017). Gaining a better understanding of the factor structure of the UWES-9 is relevant not only from a theoretical but also from a practical point of view. Specifically, confirming that the structure of the scale for which the best fit indices are obtained differs when working with different samples would imply that before using this scale it would be necessary to check its behavior in the specific population with which we intend to work; so that, depending on what is found in each particular case, appropriate decisions can be made as to whether to use and interpret the scores of each of the subscales separately, or whether it is more appropriate to use and interpret a single total score. On the other hand, the confirmation of an invariance problem makes it inappropriate to compare results obtained in studies that have worked with different samples for which different configurations have been found.

Continuing research on the psychometric behavior of the UWES-9 is especially relevant in Latin American countries where there is little scientific evidence available on this aspect (Rodríguez-Montalbán et al., 2014) and, therefore, on its applicability specifically in Ecuador. As Klassen et al. (2012) point out, "measures developed in particular settings must be carefully examined across cultural settings before they can be reliably used beyond those settings" (p. 318).

Based on the above, the present study aimed, on the one hand, to determine the most appropriate factorial structure of the UWES-9 for the measurement of the construct in its adaptation to Ecuador; and, on the other hand, to determine its predictive validity with respect to the measurement of self-assessed work performance, a variable that, as previously indicated, is the one that is most frequently predicted from work engagement.

Specifically, using Confirmatory Factor Analysis, the following measurement models were tested:

* Three correlated factors, based on studies showing high correlations between the three dimensions proposed by Schaufeli and colleagues (Balducci et al., 2010; Breevaart et al., 2012; Chaundhary et al., 2012; De Bruin & Henn, 2013; Gómez-Garbero et al., 2019; Hallberg & Schaufeli, 2006; Lovakov et al., 2017; Rodríguez-Montalbán et al., 2014; Schaufeli et al., 2006; Souza-Vázquez et al., 2017; Villotti et al., 2014; Willmer et al., 2019).
* Two correlated factors (dedication + vigor, and uptake), previously considered by Wilmer et al. (2019) and Yusoff et al. (2013).
* A unidimensional model based on the results of Fong and Ho (2015), Klassen et al. (2012) and Lovakov et al. (2017); or essentially unidimensional: (a) a classical bifactor model (a general work engagement factor and three specific factors: absorption, dedication and vigor), which allows assessing the validity of specific factor scores, controlling for the variance due to the general factor (Ondé et al. 2021); and, (b) a S-1 bifactor model or partial bifactor (De Bruin & Henn, 2013; Fong & Ho, 2015), in which one of the specific factors (i. e. reference factor) is used as the general factor (in our case, the "dedication" factor) and the rest of the specific factors are analyzed controlled by the common variance of the reference factor (Ondé et al., 2021).

**Method**

**Sample**

The study included 229 workers from private educational institutions in Guayaquil-Ecuador (73.4% women) with an average age of 36.04 years (SD = 9.38). In terms of professional category, 67.6% of the sample were highly qualified professionals and 13.1% were administrative or auxiliary personnel; 68.1% had a university degree and 16.2% had a postgraduate degree. The majority of the participants (72.1%) had more than five years of work experience, 96.8% had a permanent, indefinite contract, and 95.4% worked full time. Finally, 43.7% of the subjects had been working in their current organization for between 1 and 5 years and 28.4% had been working in the organization for more than five years.

The educational centers were selected by convenience and agreed to participate in the study. Participation was voluntary and subjects provided informed verbal consent and received information that they could withdraw at any point without consequences. All data were treated confidentially. No specific incentive was offered to participate in the study.

**Instruments**

***Work engagement***

To measure work engagement, we used the Utrech Work Engagement Scale (UWES-9) validated by Schaufeli et al. (2006), with the items in Spanish from the UWES-17 version by Salanova et al. (2000). Using a Likert scale of six intervals (1 = never/almost never to 6 = almost always/always), people were asked to indicate how often they have felt the way expressed in each of the items. As reported by Balducci et al. (2010), Chaundhary et al. (2012), Hallberg & Schaufeli (2006), Kataria et al. (2013), Klassen et al. (2012), Schaufeli et al. (2006), Villotti et al. (2014), Welfald et al. (2012), Willmer et al. (2019) and Yusoff et al. (2013), with the data obtained in the present study, the scale showed high internal consistency (α = 0.907).

***Job performance***

In the present study, job performance was defined as the behaviors, rather than the results, of the members of an organization that are relevant to the organization's goals (Campbell, 1990). To assess this construct, the Spanish translation of the English version of the Individual Work Performance Questionnaire (IWPQ-Version 1.0) by (Koopmans et al., 2016), made for the purpose of the present research, was used. The questionnaire has 18 propositions related to people's behaviour at work, and participants are asked to indicate how often they exhibited each of the behaviours in the last three months, on a six-interval Likert scale (1 = never/almost never - 6 = almost always/always). This questionnaire is composed of three scales measuring three aspects of the job performance construct, used in the present study as criterion variables, in order to assess the predictive validity of the UWES-9:

* Task performance, defined as the behaviors that contribute, directly or indirectly, to the core technical aspects of the organization (Borman & Motowidlo, 1997); that is, the competence with which the individual performs the core tasks associated with their job (Koopmans et al., 2011). A higher score on this subscale is indicative of better task performance.
* Contextual performance, defined as individual behaviors that contribute to the maintenance of the social and psychological environment in which the task performance takes place, and that go beyond what is formally prescribed as job goals (Borman & Motowidlo, 1997; Koopmans et al., 2011; Rotundo & Sackett, 2002; Viswesvaran & Ones, 2000). A higher score on this subscale is indicative of better contextual performance.
* Counterproductive behaviors, which are behaviors that are detrimental to organizational well-being (Koopmans et al., 2011; Rotundo & Sackett, 2002; Viswesvaran & Ones, 2000). A higher score on this subscale is indicative of a higher occurrence of behaviors that are counterproductive to the organization.

This instrument was chosen because a very rigorous process was followed in its development, both in terms of the literature review that underpinned the choice of items, and in the various tests conducted by its original authors (see Koopmans et al., 2011, 2014, 2014, 2016; Koopmans, Bernaards, Hildebrandt, van Buuren et al., 2013; Koopmans, Bernaards, Hildebrandt, de Vet et al., 2013) and other researchers (see Sebastian & Solana, 2016; Widyastuti & Hidayat, 2018) in order to assess their psychometric behavior. In this regard, different studies have found internal consistency indices for each of the IWPQ scales ranging from 0.76-0.87 for task performance, 0.72-0.87 for contextual performance, and 0.76-0.89 for counterproductive behaviors (Koopmans et al., 2016; Ramos-Villagrasa et al., 2019; Gabini & Salessi, 2016; Widyastuti & Hidayat, 2018). The different scales of the Spanish adaptation of the IWPQ-Version 1.0 used in the present research also showed adequate internal consistency (Task performance: α = 0.733; Contextual performance: α = 0.793; Counterproductive behaviours: α = 0.771).

**Data analysis**

We conducted Confirmatory Factor Analysis (CFA) to fit different measurement models: (a) three correlated factors (3F), (b) two correlated factors (2F), and (c) one factor, either unidimensional (1F) or essentially unidimensional.

For the CFA we used the Lavaan package (Rosseel, 2014) in the R Project for Statistical Computing (version 4.1.0), and the diagonally weighted least squares estimator (DWLS) using a polychoric correlation matrix. This estimator is recommended for analyzing samples with a small-moderate number of observations with ordinal data, such as that yielded by Likert-type items (Asún et al., 2016; Jöreskog & Sörbom, 1989). For the CFA model fit evaluation, we used the χ2 tests, the Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Comparative Fit Index (CFI), and Tucker-Lewis Index (TLI). We used the following criteria for evaluating model fit: RMSEA < 0.08, SRMR ≤ 0.08, CFI and TLI ≥ 0.95 (Brown, 2015).

SPSS 25.0 was used for the descriptive analysis of the UWES-9 items and the regression analyses aimed at assessing the predictive validity of the UWES-9.

**Results**

**Preliminary Analysis**

In the case of work engagement, item means ranged from 5.07 to 5.41, with standard deviations between 0.815 and 1.113 (Table 1), indicating that, in general terms, participants reported high levels of work engagement. In this sense, the distributions presented a negative skewness indicating that, in all items, the scores tended to cluster towards the high values of the scale, showing a moderate homogeneity. In addition, the distribution of the data was leptokurtic for all items except item 2 (Table 1).

**CFA Model Fit**

 The bifactor models (classical bifactor and S-1 bifactor) did not converge; the rest of the models, as shown in Table 3, showed adequate goodness of fit in the CFI, TLI and SRMR indices; but, not in RMSEA and especially χ2 (see Table 3 and Figure 1), which are very sensitive to small problems of correct model specification.

The ANOVA results showed that, comparing the three models, the three-factor model fits significantly better than the two-factor model, and the two-factor model fits significantly better than the one-factor model (Table 4).

Table 4.

*Results Obtained in ANOVA Performed to Determine Differences in the Goodness of Fit of Models 3F, 2F and 1F*

 The reason for the three models to show a good fit is the strong correlation observed between the factors considered (vigor [VIG]-dedication [DED] = 0.926; vigor-absorption [ABS] = 0.916; dedication-absorption = 0.946), which allows us to propose the unidimensional model as the most parsimonious option with the advantage associated with this type of structure, whether unidimensional or essentially unidimensional (i.e. bifactor), which is to be able to use the overall score of the test, instead of having to apply it as a multidimensional instrument (Trizano-Hermosilla et al., 2021).

 However, the fact that χ2 is significant, without the circumstance of a large sample size, is indicative of a certain degree of misspecification. In fact, Balducci et al. (2010), Klassen et al. (2012) and Lovakov et al. (2017) observed correlation between errors due to content overlap between two items of the Vigor dimension: V1 (At my workplace, I feel full of energy) and V2 (At my job, I feel strong and vigorous). In addition,

Klassen et al. (2012), in their Australian and Canadian samples, and Lovakov et al. (2017) found content overlap between two items of the Absorption dimension: A8 (I am immersed in my work) and A9 (I get completely lost in my work). Finally, Balducci et al. (2010) observed correlations between the errors of items: V2 and D3 (I am enthusiastic about my job), and Lovakov et al. (2017) found correlation between the errors of items: V5 (When I get up in the morning, I feel like going to work) and A6 (I am happy being absorbed in my work).

In these cases, the analyses of the modification indexes indicated the loss of gain due to the non-inclusion of these correlations.

 In the same way as in the research by Balducci et al. (2010), Klassen et al. (2012) and Lovakov et al. (2017), in the present study, correlations were found to exist between errors found for the following pairs of items: V1 and V2, and A8 and A9. Thus, some general content overlap between pairs of items is indicated. Additionally, our results showed the existence of a significant error correlation between item D7 of Dedication (I am proud of the work that I do) and A8 of Absorption (I am immersed in my work). Through including these corrections in the models (3F&CE, 2F&CE, 1F&CE), it was observed that the goodness of fit was now excellent in all the indices, including the χ2 test (see Table 5).

In order to determine in terms of goodness of fit χ2 which model was the best, an ANOVA was applied (see Table 6) in which no significant differences were observed between the three models considered. Thus, with the imperative to consider the parsimony criterion, this would lead us to choose the unidimensional model as the most appropriate.

Table 6.

*ANOVA on the differences in the Goodness-of-Fit of the 3F&CE, 2F&CE and 1F&CE Models*

**Predictive Validity of the UWES-9**

As the factor structure of the UWES-9 conforms to a unidimensional model, using the overall score of the instrument to assess its predictive validity is warranted. In turn, as the instrument distinguishes three content domains (vigor, dedication, and absorption), the predictive ability of the subscales was also assessed.

Self-assessed job performance was used as a criterion, measured with the Spanish adaptation of the IWPQ-Version 1.0, which reports three key factors: task performance, contextual performance and counterproductive behaviour. Simple regression analyses were performed, using the overall UWES-9 score as a predictor.

According to the results, the overall score or WE of the UWES-9 significantly predicts the task and the contextual performance. Therefore, to the extent that workers show higher levels of WE, they consider their performance to be better, both in the core tasks associated with their work and in those that contribute to good organizational performance even when they are not formally associated with the work objectives. Furthermore, it was found that the WE explained a higher percentage (33.4%) of the variance of contextual performance than of the variance of task performance (23.8%). However, the WE did not significantly predict the counterproductive behavior (Table 7).

Table 7.

*Results of Simple Regressions Performed to Evaluate the Predictive Power of WE as a Unidimensional Measure*.

To determine the predictive power of each of the theoretical dimensions of the UWES-9, a multiple regression analysis was performed for each of the aspects of job performance used as criterion variables. The results showed the relevance of considering the scores in each of the UWES-9 subscales, since they provide different predictive information that may be useful depending on the purpose for which the WE is evaluated.

In the case of task performance, although the model significantly predicted this variable (R = 0.489; R2 = 0.239; F[1, 217] = 22.702, *p* = 0.000), the only dimension of the UWES-9 that significantly predicted task performance was "vigor" (Table 8); thus, participants considered their performance on tasks directly associated with their work to be higher as they experienced higher levels of energy and physical and mental stamina while working.

As for contextual performance, this variable was also significantly predicted by the model (R = 0.584; R2 = 0.341; F[1, 217] = 33.446, *p* = 0.000), and both "vigor" and "absorption" significantly predicted the contextual performance, with vigor being the best predictor (Table 8). Thus, people indicated that they performed more frequently behaviors that, although not explicitly associated with their work objectives, contribute to the adequate performance of the organization, as they experienced greater vigor and felt more focused and absorbed in their work.

Finally, with regard to counterproductive behaviours, this variable was significantly predicted by the model (R = 0.199; R2 = 0.040; F[1, 215] = 2.965, *p* = 0.033); but the only relevant dimension of the UWES-9 was "dedication" (Table 8); so that, to the extent that people reported feeling more enthusiastic, inspired and proud of the work they do, they were less likely to exhibit behaviours that were detrimental to the well-being of the organization. The findings on counterproductive behaviors help to explain why, when the overall scale score is used as a predictor, it does not significantly predict the counterproductive behavior: the sum score cancels out because the counterproductive behavior is negatively related to the dedication factor, which is to be expected from the theory; but, it tends to be positively related to the absorption factor, which is contrary to what is expected from the theory, since, in principle, what would be expected is that the frequency of occurrence of counterproductive behaviors for the organization is lower in workers who report feeling more focused and absorbed in their work.

Table 8.

*Results of the Multiple Regressions performed to Evaluate the Predictive Power of the WE Dimensions.*

**Discussion**

The purpose of this study was to determine which factor structure of the UWES-9 is the most appropriate for the measurement of the construct in Ecuador, by comparing different models derived from work engagement (WE) theory and introducing empirical evidence, as well as to evaluate its predictive validity with respect to the measurement of self-assessed work performance.

Regarding the factor structure of the UWES-9, first of all, our findings do not support the proposals of authors such as De Bruin and Henn (2013) and Fong and Ho (2015) regarding the possible suitability of a partial bifactor model, since the S-1 bifactor model with the dedication factor as the reference factor did not cover; and the classical bifactor model did not converge either, probably as a consequence of the good goodness of fit of the unidimensional model. Surely in samples in which the correlation between factors is lower these models that propose essential unidimensionality can be a good alternative to the unidimensional model.

Secondly, the results of the CFA conducted for the three-factor correlated, two-factor correlated (dedication+vigor, and absorption) and unifactorial models, evidenced that all three models showed CFI, TLI and SRMR indices indicating good fit; but, this was not the case for RMSEA and the χ2 test. This is consistent with that reported by Hallberg and Schaufeli (2006), Kataria et al. (2013), Klassen et al. (2012), Sousa-Vasquez et al. (2017), Villoti et al. (2014), Welfald et al. (2012) and Willmer et al. (2019). However, in the present study the models did differ significantly in terms of goodness of fit, with the 3F model being significantly superior to the other two. This superiority of the three-factor model coincides with the findings of authors such as Balducci et al. (2010), Breevaart et al. (2012), Gómez-Garbero et al. (2019), Lovakov et al. (2017), Rodríguez-Montalbán et al. (2014), Schaufeli et al. (2006), and Yusoff et al. (2013).

However, this level of analysis overlooks the fact that both in this and in other investigations a possible problem of misspecification has been detected, as the scale includes some items whose contents overlap. Specifically, in the Ecuadorian sample we have confirmed the content overlaps observed by Balducci et al. (2010), Klassen et al. (2012) and Lovakov et al. (2017) between items V1 and V2, and by Klassen et al. (2012) and Lovakov et al. (2017) between items A8 and A9. Additionally, in the present study we detected a new, previously unreported overlap between items D7 and A8 that should be investigated in future studies to determine whether this is a problem specific to the sample we worked with in the present study, or whether there is truly a relationship between these two indicators that is generalizable to other samples and cultures.

 When CFA is performed in order to correct for this possible redundancy problem, the three models show excellent fit indices, with no statistically significant differences between them. On this point our results differ from what is reported by other authors in the sense that, in our case, by allowing for correlations between errors, the models were equivalent in terms of goodness of fit even considering the χ2 test; whereas, in all the studies consulted the authors conclude on the superiority or not of one model over another without considering the value of χ2 and in all previous research the χ2 was statistically significant. In fact, Fong and Ho (2015) point out that one of the methodological problems frequently encountered in studies on the UWES-9 is that researchers ignore significant χ2 values claiming that this test is very sensitive to detect model misfit when working with large samples. However, in actuality, the value of this fit index is usually ignored even in studies with small-moderate sample sizes.

The findings regarding the goodness of fit of the models by allowing correlations between errors would lead us to think that, since more complex structures do not represent an advantage over simple ones, parsimony would justify the use of the UWES-9 as a unidimensional instrument of the WE. This unidimensionality has some advantages from a metric point of view, such as being able to use the overall test score, which, according to Klassen et al. (2012) and Schaufeli et al. (2006), may be preferable for the use of the UWES-9 across settings.

Nonetheless, this option clearly clashes with the theoretical conceptualization of WE as a multidimensional construct. The creators of the UWES-9 did not intend to construct a unidimensional instrument when developing the scale, but rather to reduce the number of items in order to maintain the distinction between the three dimensions. There is thus a conflict between the principle of parsimony and the theoretical proposal, the clarification of which requires a brief reflection on the theoretical model and the goodness of fit of the models.

A common mistake is to confuse unidimensionality with homogeneity of construct measurement. The fact that the unidimensional model shows a goodness of fit equivalent to that of a three-factor model does not imply that the dimensions do not exist. Rather, it simply implies that from the point of view of empirical fit it is not possible to differentiate the response to the items of each dimension. Thus, our results are evidence that, in the sample analyzed, the scale does not allow differentiation between the factors. This only indicates that, in different samples, one and other theoretical components operate and relate to each other in different ways; in some, as was our case, they are so close that it is not possible to differentiate them, in other samples two of them can be observed or the three elements are so differentiated that the best-fitting model is the two-factor or "uncorrelated" factor model.

However, the dimensions exist from the moment that the researcher has enunciated items for the evaluation of these dimensions. Thus, for example, item 1 "At my workplace, I feel full of energy" (Vigor), is an indicator that tracks a content domain clearly different from item 7 "I am proud of the work I do" (Dedication), and these in turn differ from item 8 "I am immersed in my work" (Absorption). One should also note that vigor, dedication and absorption go hand in hand, so that a person who feels more involved and identified with their work, in turn, puts more effort into it and feels more absorbed when doing it.

In the educational setting, where professionals tend to be highly engaged, it is not surprising that the three dimensions are strongly correlated, and that the most parsimonious model is the unidimensional one, which should not be confused with a simple measure or underrepresentation of the content domain. As a unidimensional construct, our results show that, also in Ecuador, WE significantly predicts the task and self-evaluated contextual performance, which corroborates the reports of other researchers (Afacan-Findikli, 2015; Breevaart et al., 2015; Christian et al., 2011; Dalal et al., 2012; Gutermann et al., 2017; Kim et al., 2013; Lorente et al., 2014; Lupano et al., 2018; Motyka, 2018; Pongton & Suntrayuth, 2019; Qodariah et al., 2019; Rich et al., 2010). Furthermore, it was found that the WE explained a higher percentage of the variance of contextual performance in relation to variance of task performance. According to Borman and Motowidlo (1997), this may be due to one of the differences between task and contextual performance is that the contextual performance is more likely to be determined by relatively stable characteristics of individuals, such as the WE; whereas, the antecedents of the task performance rather involve cognitive abilities. However, contrary to what would be expected, the work engagement did not allow predicting the counterproductive behaviors.

This does not mean that the WE dimensions are irrelevant; in fact, they are irrelevant not only from a theoretical point of view, but also because they allow predicting different aspects of job performance. Specifically, our results showed that task performance was significantly predicted only by vigor, the counterproductive behaviors were predicted only by dedication, and the contextual performance was predicted by both vigor and absorption, with vigor being the best predictor. Following these results on the predictive validity of the UWES-9, the recommendation would be to use the trifactorial model, since the unidimensional model leads to the loss of relevant information.

**Conclusions**

Based on the above, we conclude that, with regard to the structure of the UWES-9, our results are consistent with previous evidence that it is inappropriate to automatically generalize to any context the notion of work engagement as a construct consisting of the three dimensions proposed by Schaufeli and colleagues. Thus, there is a problem in terms of the magnitude of the correlations between the factors observed in different samples, which is aggravated by minor problems of content overlap between some items, as revealed by observed inter-error correlations.

Consequently, given the implications of accepting one or the other models (e.g., whether to use the total scale score or to work as if it were a multidimensional instrument with the scores of each subscale), we agree with authors such as Klassen et al. (2012) in asserting that conducting the CFA in each case is key for the applied researcher in order to understand well how the measurement of the construct operates in a given sample and to carry out the relevant analyses.

Any decision to accept one structure or the other should not be guided solely by goodness of fit or the principle of parsimony. Rather, the key is to reconcile theory with empirical evidence. In the search for this reconciliation, at least in the case of Ecuador, although it is possible to use the total score since there is an acceptable unidimensional structure, the three-factor proposal is superior from a substantive and predictive point of view.

**Limitations**

One of the main limitations of the present study is the use of a simple sample of a specific occupation, with predominately female participants. It is probable that if multiple samples with different characteristics had been collected, it would have been possible to reproduce all the variability of models proposed in the literature. Such an exercise would be desirable to validate our proposal that any empirical application due to the characteristics of the participants themselves will lead to small distortions and that these should be taken into account for the correct use of the instrument scores.