Shorter may be better: metric properties of the uwes-3 to assess engagement in mexican university professors

Arturo Juárez-García *¹; César Merino-Soto²; Daniela Villamar Sánchez³

¹Centro de Investigación Transdisciplinar en Psicología, Universidad Autónoma del Estado de Morelos, México https://orcid.org/0000-0003-3264-679X arturojuarezg@hotmail.com ²Instituto de Investigación de Psicología, Universidad de San Martín de Porres, Lima, Perú, https://orcid.org/0000-0002-1407-8306 sikayax@yahoo.com.ar ³Instituto de Ciencias de la Educación, Universidad Autónoma del Estado de Morelos, México https://orcid.org/0000-0002-5843-9628 danielavillamarsanchez@gmail.com

Abstract

Introduction: Work engagement is a relevant positive psychological effect in the literature, and validity studies of the ultra-brief version of the Utrecht Work Engagement Scale are practically nonexistent in Latin America. **Objective:** to identify the psychometric properties of the ultra-brief version of the Utrecht Work Engagement Scale (UWES-3) and to analyze its equivalence with the UWES-9 version in a sample of teachers from a public university in Mexico. **Method:** the final sample consisted of N=247 university teachers, who answered the 9-item and 3-item versions of the UWES and the Burnout Evaluation Questionnaire for Education Personnel (CESQT-PE) to estimate its discriminant validity. **Results:** The results confirmed very high correlations between both versions of the UWES (3-9) (r=>.82-.93), a unidimensional structure of the UWES- 3 with higher factor loadings (>.75), satisfactory internal consistency values (>.70) and high and similar correlations in the expected direction between both versions and the CESQT-PE (q=.02-.08). **Conclusions:** We conclude that there is equivalence of the UWES-3 and the UWES-9. The practical benefits of the ultra-brief version for the measurement of work engagement in university teachers are discussed.
Keywords: Engagement; UWES; University teachers; Measurement Scales.

Más breve puede ser mejor: Propiedades métricas de la UWES – 3 ítems para evaluar "engagement" en profesores universitarios mexicanos

Resumen

Introducción: el *engagement* laboral es un efecto psicológico positivo relevante en la literatura, y prácticamente son inexistentes estudios de validez de la versión ultra-breve de la Utrecht Work Engagement Scale en Latinoamérica. **Objetivo:** identificar las propiedades psicométricas de la versión ultra-breve de la Utrecht *Work Engagement Scale* (UWES-3) y analizar su equivalencia con la versión UWES-9 en una muestra de docentes de una Universidad Pública de México. **Método:** la muestra final fue de N=247 profesores universitarios, quienes contestaron la UWES en sus versiones de 9 y 3 ítems, así como el Cuestionario de Evaluación de Síndrome de Quemarse por el Trabajo para personal de edu¬cación (CESQT-PE) con el propósito de estimar su validez discriminante. **Resultados:** los resultados confirmaron correlaciones muy altas entre ambas versiones

*Correspondencia:
Arturo Juárez García
arturojuarezg@hotmail.com

del UWES (3-9) (*r*=>.82-.93), una estructura unidimensional del UWES- 3 con mayores cargas factoriales (>.75), valores de consistencia interna satisfactorios (>.70) y correlaciones altas en la dirección esperada y similares con el CESQT-PE (q=.02-.08). **Conclusiones:** se concluye que existe equivalencia del UWES-3 y el UWES-9 y se discuten los beneficios prácticos de la versión ultra-breve para la medición del *engagement* laboral en docentes universitarios..

Palabras clave: entusiasmo laboral; UWES; Docentes universitarios; Escalas de medición

Introduction

Work engagement is perhaps the most popular positively focused work-related psychological effect in recent decades. Although attempts have been made to translate it into Spanish using concepts such as *entusiasmo laboral* (enthusiasm towards the job) (Juárez-García, 2015), the English term is still the most common, both in consulting and academic contexts, perhaps so as not to confuse it with other psychological phenomena. Maslach and Leiter (1997) positioned it just as a direct opposite to burnout syndrome, that is, feeling energetic as opposed to the symptom of emotional exhaustion; being psychologically involved in the work activity as opposed to the depersonalization of burnout; and, finally, feeling professionally self effective as opposed to the absence of such certainty about capabilities when there is psychological burnout. However, this paradigm suffers from an approach restricted to two mutually exclusive opposites, where the basic logic is that low levels of burnout automatically represent high levels of engagement, and therefore both can be measured with the same instrument (Maslach Burnout Inventory [MBI]) (Maslach et al., 1986), which has evidently been criticized (González-Roma et al., 2006).

According to Schaufeli et al., (2002), work engagement is a phenomenon negatively correlated to burnout, but at the same time independent. It is defined as a positive motivational effect that is derived from work and is characterized by three components: 1) vigor, which is characterized by high levels of energy and

mental resilience while working and willingness to strive and persist in difficult situations at work; 2) dedication, which refers to being strongly motivated at work and experiencing a sense of meaningfulness, inspiration, pride, and challenge; and 3) absorption, which refers to being completely focused and immersed in work, wi*th t*he feeling that time flies by in a kind of total engagement with work (Salanova & Schaufeli, 2009). Engagement's relationship with motivation, satisfaction, happiness, initiative, and good work performance has been demonstrated among various favorable individual and organizational effects in different contexts (Salanova & Schaufeli, 2009). In fact, engagement has even been associated with the productive efficiency of countries, the gross domestic product, the presence of integrity, and the absence of corruption, as well as with cultures with high individualism, high enjoyment of life, and low power distance (Schaufeli, 2018).

Work engagement in the teaching practice has seen little research; however, positive relationships have already been demonstrated between this phenomenon and teachers' satisfaction (Han et al. , 2020), the presence of more positive emotions at work (Burić & Macuka, 2018), greater effectiveness in their work activity (Wyatt & Dikilitaş, 2016), and even the improvement of cognitive performance of their students (Zhu, 2001), among many other positive consequences.

A turning point in the development of the construct came with the work of Schaufeli et al. (2002), who not only demonstrated that the construct was independent of burnout (Gonzalez-Roma et al., 2006), but also defined it and conceptually differentiated it from other constructs based on empirical evidence. In addition, they proposed a measure with satisfactory psychometric properties in different samples and regions: the Utrecht Work Engagement Scale (UWES) (Salanova et al., 2000; Schaufeli et al., 2002; Schaufeli, 2017).

The UWES scale has been validated in several languages in more than 30 countries on five continents (Schaufeli & Bakker, 2010). Its initial 24 item version was validated in a proposed 17 item functional scale, with the expected trifactorial structure and internal consistency indices (α)

above .70 in student and worker populations (Schaufeli et al., 2002). Subsequently, Salanova et al., (2000) tested a 15 item version of the UWES in a sample of workers using information technologies, including five questions for vigor, five for dedication, and five for absorption on a seven option Likert response scale ranging from "0 = Never" to "6 = Every day". The results confirmed the three-dimensional structure and internal consistency indices (α) higher than .73 in all subscales, as well as independence with the burnout construct and statistically significant relationships in the expected directions with organizational indicators such as organizational commitment, job satisfaction, or work enjoyment.

In a next step, Schaufeli et al., (2006) validated a 9-item version of the UWES (3 for each component) in 10 European countries of different occupational groups. The internal consistency was satisfactory (α > .70), and, although the trifactorial structure was tested, the best indices were for a unidimensional structure. Given the high correlations between the three components, the authors suggested the use of the total score in a single global measure of engagement.

More recently, Schaufeli et al., (2019) validated an ultra brief version of only three items (UWES-3) in samples from several countries. It maintained good psychometric indices, shared up to 92% of variance with the 9-item version (UWES-9), and presented no differences in correlation patterns with other variables. The authors suggest the use of the UWES-3 in future studies and its incorporation in national surveys, given all the practical advantages it represents.

Although there are some validations of the work engagement scale (UWES) in Latin America in samples of teachers (Flores et al., 2015; Álvarez Garzón & Peña Fuentes, 2019; Martins & Mendonça, 2019), these have been conducted with the 17-, 15-, or 9-item versions, and only one very recent publication was identified that validated the ultra brief UWES-3 version in Peruvian students. This is the Peruvian study by Domínguez Lara et al., (2021), who concluded that the psychometric properties of this brief version did not differ from the other versions of the UWES. Nevertheless, validation studies of this ultra brief version in

populations of academic workers in Latin American populations are still pending, particularly in Mexico, where there are no studies in this regard. The practical and research implications of having an ultra brief measure for screening the engagement phenomenon in teachers represent multiple advantages associated with simple, rapid detection, and with the best possible effectiveness cost and efficiency time ratios in the evaluation of this phenomenon in teachers.

Given the above, the aim of the present study is to analyze the psychometric properties of the UWES-3 ultra brief version and its equivalence with the UWES-9 in a sample of Mexican professors.

Method

Participants

Considering a total population of 1,825 professors and researchers in the thematic and regional centers of Universidad de Guadalajara, an email was sent to all of them with the invitation to participate and a link to access the survey through the SurveyMonkey® platform. To access the questionnaire, it was necessary to sign the letter of consent in agreement. A total of 259 academics agreed to participate, so the response rate was 14.19%. The inclusion criteria were the following: to be carrying out research and teaching activities, to be an active worker at the university at the time of the study, to have a full time contract, to agree to participate in the study, and to sign the informed consent letter. The exclusion criteria were the following: being on sabbatical, to have a contract other than full-time, and to not be participating in any stimulus program. The elimination criterion was not answering the battery of instruments or answering it partially. Finally, according to these criteria, a sample of 247 academics was obtained.

The final sample was distributed as follows: 122 men (49.4%) and 125 women (50.6%), with an average age of 49.28 (standard deviation = 9.677). Regarding marital status, 149 were married (60.3%), 53 were single (21.5%), 23 cohabitated (9.3%), 18 were divorced (7.3%), and 4 were widowed (1.6%). In terms of highest level of studies, 38 had a

master's degree (15.4%), 179 had a PhD (72.5%), and 30 had a postdoctoral degree (12.1%). Regarding the university center of affiliation, 205 belong to a Thematic Center (82.9%) and 42, to a Regional Center (17.1%).

As to their official titles, 240 participants indicated that they were professor-researchers (97.2%), and 7 indicated that they were academic technicians (2.8%). However, all of them were engaged in research and teaching activities. The average length of service in the institution was 19.32 years (standard deviation = 10.983), although the average in the current position was 8.81 years (standard deviation = 7.832).

Ethical Considerations

An informed consent letter was elaborated. This document requested the voluntary participation of academics, guaranteed the confidentiality of their identity, and gave them the freedom to leave the survey at any time they wished. There were no personal identifiers or invasive procedures, and all the recommendations of the Helsinki declaration (World Medical Association [WMA], 2013) were followed.

Instrumentos

Following the same logic of the available studies on the validity and equivalence of the short version of the UWES-3, the 9-item version was selected as a concurrent criterion because it is the improved version of previous ones and the one with better psychometric capabilities than the 17- or 15-item versions (Schaufeli et al., 2006). Likewise, as an external criterion to assess the equivalence patterns of correlations between both versions, age, and sex were considered, as well as a measure of burnout or emotional exhaustion with which a divergent type of validity or correlation with the UWES measures is expected, as it happened in initial studies. In this study, the CESQT was used, an instrument that denominates burnout as work burnout syndrome. These instruments are broken down below:

Utrech Work Engagement Scale (UWES) in its 9- and 3 Item Versions (Schaufeli et al., 2017**).** It is an engagement assessment measure within the same model in which its previous versions (i.e., UWES-15 and UWES-9) were created. The ultra brief UWES-

3 version is composed of three items representing each of the original components referring to feeling energy (item 1, vigor), enthusiasm (item 2, dedication), and immersion (item 8, absorption). Its response options are ordinally scaled from 0 (never-at no time) to 6 (always-every day). The internal consistency reliability (α) in the original study ranged from .77 to .85.

Burnout Syndrome Evaluation Questionnaire for Education Personnel (CEQST-PE). Contains 20 items distributed in four dimensions (Gil-Monte, 2011). Job Enthusiasm (1) implies the individual's desire to achieve work goals because they are a source of personal satisfaction. Psychic exhaustion (2) refers to emotional and physical exhaustion due to the fact that, at work, one has to deal daily with people who pose or cause problems. Indolence (3) represents the presence of negative attitudes of indifference and cynicism towards the organization's clients (patients, students, etc.). Guilt (4) involves feelings of guilt that the person has for the behavior and negative attitudes developed at work, especially towards people with whom work relationships are established. The response options include a five-grade frequency scale ranging from 0 (never) to 4 (very frequently; every day) (Gil-Monte & Noyola, 2011). The CESQT-PE has shown satisfactory validity and reliability in Mexican and Latin-American samples (Gil-Monte et al., 2009) and with superior efficacy to the Maslach Burnout Inventory Human Services Survey (MBI-HSS) (Juárez-García, 2015; Calderón de la Cruz et al., 2020), so the scale and its dimensions are a suitable criterion in the discriminant validity of the UWES-3.

Procedure

Analysis. It consisted of two procedures: a preliminary initial one to prevent the effect of unintended response distortions (Wetzel et al., 2016) and to describe the responses to the items, and the second one focused on the main objective of the study.

Preliminary analysis. In the first one, the data were scanned to detect possible response biases expressed by multivariate outliers (Meade & Craig, 2012). Considering its efficiency (Meade & Craig, 2012), the D2 distance (Mahalanobis, 1936) was used with the normtest program (DeCarlo,

1997). Participants detected as outliers were removed from the database. Then, at the item level, descriptive, distributional, and correlational statistics (equality of inter-item correlations; Steiger, 1980) were obtained in the R psych (Revelle, 2020) and R companion (Mangiafico, 2021) programs.

Psychometric analysis. For the main objective, the analysis of psychometric properties was sequentially performed by assessing dimensionality (i.e., number of latent factors), scalability of scores and items observed, structural characteristics, equivalence between groups, and correlation with other variables.

Dimensionality. Because there is discrepancy between the most appropriate construct representation with the UWES, the number of latent variables was first explored by assessing the number of latent factors. A consensus procedure was implemented among nine available factor extraction methods (e.g., parallel analysis, optimal coordinates, MAP-Velicer, etc.) using the psycho program (Makowski, 2018). The Data Comparison method CD (Ruscio & Roche, 2012) and RGenData program (Ruscio, 2018) were also used because it tends to show adequate accuracy (Auerswald & Moshagen, 2019).

Structural properties. Two approaches were applied: a nonparametric one, appropriate for measures with few items (Van Schuur, 2003), and a parametric one, conducted in that order. Regarding the nonparametric model, the predominant dimensionality found in the previous analysis was corroborated by means of the Mokken Scale Analysis (MSA) (Mokken, 1971). As a proxy for the number of latent dimensions (Straat et al., 2013), the Automated Item Selection Procedure (AISP) (Straat et al., 2013) algorithm was applied to select scalable items within homogeneous groups and identify the number of scales. The range of .30 through .70, in steps of .10 units of the scalability coefficient H was used for the detection of homogeneous groups of items. Once this was completed, the next step corroborated the properties of the observed scores, which are precursors for more complex parametric modeling (e.g., SEM or IRT-based). Thus, compliance with the monotonic homogeneity model (MHM) (Sijtsma

& Van der Ark, 2017) was verified by assessing a) scalability with coefficient H (Loevinger, 1948) for score and items (Hi), where coefficients greater than .40 are acceptable (Sijtsma & Van der Ark, 2017); b) local independence by means of coefficients W(1) and W(^{3}) for inter-item relationships and W(^{2}) for each item (Straat et al., 2016); and, finally, c) item monotonicity (i.e., the incremental monotonic relationship between items and scale score) by means of item characteristic curve plots. Finally, from the MSA, reliability was estimated with the MS coefficient. For the MSA, the Mokken program (Van der Ark, 2012) was used.

In the second approach, using linear parametric modeling, structural equation modeling (SEM) was used with the WLSMV estimator for categorical variables and practical fit indices, such as CFI (≥ 95), TLI (≥ 95), and SRMR (≤ .05). Three models were evaluated: correlated and unidimensional dimensions in the UWES-9, and unidimensional in the UWES-3. The lavaan (Rosseel, 2012) and semTools (Jorgensen et al., 2020) programs were used.

Reliability. The internal consistency reliability was examined using the MS coefficient (Molenaar & Sijtsma, 1984, 1988) through Mokken Scaling Analysis modeling; the alpha (α) and omega (ω) coefficients were estimated through linear modeling, with confidence intervals at 95%.

Equivalence. Equivalence between UWES-9 and UWES-3 versions was assessed by means of a) the adjusted correlation between their direct scores (Levy, 1967) and the AC1 coefficient (Gwet, 2008, 2019), for agreement in the classification of participants into quartiles. Finally, construct validity equivalence was done by comparing dependent correlations (Zou, 2007) between UWES-9 and UWES-3 with CESQT scales and demographic variables (sex and age). The cocor program (Diedenhofen & Musch, 2015) was used.

Results

Preliminary Analysis
Item Descriptive Information. The UWES response trend (Table 1) was between options 3 (*regularly: a few times a month*) and 4 (*frequently: once a week*),

while the first option (*sporadically: a few times a year or less*) was predominantly low and there was a poor response frequency; this was more evident in items 7 and 8, where there were no responses in the first option. Differences in the average ranking of each item were statistically significant (Friedman χ^2 = 278, df = 8, p < .001), with Kendall - W effect size = .523 (95% IC = .514, .262). Regarding inter-item correlations (not shown in Table 1), correlational identity was not maintained (χ^2 = 2601.53, df = 36 [p < .01]), indicating that there are differences in magnitude between inter-item correlations.

Detection of response bias. Using the Mahalanobis *D2* procedure, the critical value established in the F distribution (9, 249) was D2 > 30.67 (.05 level with Bonferroni adjustment); thus, 6 participants with *D2* values between 31.41 and 52.97 were removed from the database. Visual examination of the response pattern of these subjects showed response inconsistency on all items of the UWES.

Dimensionality

The dimensionality of the UWES-9 on a single factor was supported by five methods (55.5%; optimal coordinates, acceleration factor, parallel analysis, MAP-Velicer, and VSS complexity 1), while three methods (33.3%) supported two factors, and only one method (BIC adjusted for sample size) indicated three factors. The CD method indicated two possible factors. The greatest convergence occurred for one-dimensionality with the most recommended methods in research practice (Auerswald & Moshagen, 2019), and therefore the factor analysis was oriented toward estimating the parameters of a general unidimensional model of the 9 item version. From another approach, with the nonparametric MSA approach (see next section), we proceeded to verify one-dimensionality.

Mokken Scale Analysis (MSA)

Scale Identification. First, the AISP procedure showed more clearly that a single scale can be

Tabla 1.

Frecuencia de opciones de respuesta del UWES-9 y resultados correlacionales

	Correlaciones Spearman					Frecuencia de respuesta				M	DE
	Sexo	Il. Trab	Indol.	Desg.	Culpa	1	2	3	4		
uwes1 a	-.114	.455**	-.193**	-.485**	-.220**	2	44	137	70	3.09	.690
uwes2	-.120	.361**	-.177**	-.393**	-.239**	1	37	152	63	3.09	.635
uwes5	-.140*	.514**	-.185**	-.366**	-.144*	3	26	136	88	3.22	.671
uwes3 a	-.139*	.565**	-.240**	-.363**	-.183**	1	16	139	97	3.31	.605
uwes4	-.098	.565**	-.307**	-.328**	-.266**	1	13	122	117	3.40	.607
uwes7	-.018	.410**	-.258**	-.174**	-.174**	0	5	90	158	3.60	.529
uwes6	-.028	.475**	-.189**	-.202**	-.200**	1	22	102	128	3.41	.664
uwes8 a	-.008	.407**	-.188**	-.073	-.101	0	7	125	121	3.45	.551
uwes9	.058	.301**	-.103	-.045	-.018	6	47	123	77	3.07	.763

Nota. Il. Trab.: Ilusión por el Trabajo. Indol.: Indolencia. Desg.: Desgaste. a Ítems de la versión ultra-breve UWES-3

obtained in the H range between.30 to .50 (Table 2, left side). At the highest level (H > .50), no three-scale configuration was identified, which converges with the parametric analysis in the previous paragraph, in that the items converge on a single dimension of the UWES 9.

Scalability. The H coefficient for the total score (Table 2, middle part) was high (H = .554, s.e. = .033, CI 95%: .489, .618); in males, scalability was moderately higher (H = .604, s.e. = .044, CI 95% = .517, .690) compared to females (H = .492, s.e. = .048, CI 95% = .397, .586). The inter-item scalability (Hij) ranged from .869 (s.e. = .059) and .029 (s.e. = .098), with item 9 showing the lowest coefficients with the rest of the items. Finally, the scalability for each item (Hi) was higher than .540, except for item 9. All coefficients H were statistically significant (z > 2.50).

Local Dependence. $W^{(2)}$ ranged from 27.82 to 41.80, $W^{(1)}$ ranged from .370 to 7.565, and $W^{(3)}$ ranged from 8.33 to 9.248; none were identified as substantial evidence of local dependence between items.

Monotonicity. Figure 1 shows the incremental pattern of response options and item scores. The least differentiated curves are those corresponding to the first response options, and all items maintain a constant and practically linear increase.

Linear Modeling (SEM)

In the UWES-9 version (Table 3), the three-factor correlated model fitted satisfactorily: WLSMV-χ^2 = 122.636, df = 2 (*p* < .01); CFI = .989, TLI = .984, SRMR = .076; however, the interfactor correlations were high, suggesting lack of conceptual discrimination between them. The unidimensional model was also satisfactory, WLSMV-χ^2 = 316.665, df = 27 (*p* < .01); CFI = .968, TLI = .958, SRMR = .122, although slightly low. The UWES-3 fit was completely satisfactory: WLSMV-χ^2 = 638.132, df =3 (*p* < .01); CFI = 1.000, TLI = 1.000, SRMR = .000. Factor loadings in all models were high (> .50) and statistically significant.

Item 3 in the UWES-3 model showed loading greater than 1.0, but this was not necessarily a Heywood case due to its magnitude. No modifications to the unidimensional UWES-9 model were detected, and they were not explored.

It is also observed that the items corresponding to the UWES-3 consistently had the highest factor loadings compared to the rest of the items in the tridimensional and unidimensional models (Table 3). The thresholds (the separation of the options in the latent continuum of the UWES 3 score) were irregular in their separation, and item 8 showed that the last response option was not functional due to the absence of responses, and the threshold was not estimated.

Tabla 2.
Escalamiento Mokken: Procedimiento automatizado de selección de ítems (AISP)

	AISP (normal method)					Scaling (Hi)			Monotonic homogeneity (MHM)		
	.30	.40	.50	.60	.70	Total	Hom	Muj	#ac	vi	CRIT
uwes1[a]	1	1	1	1	1	.542	.629	.444	7	0	0
uwes2	1	1	1	1	1	.542	.581	.484	7	0	0
uwes5	1	1	1	1	0	.576	.611	.529	7	0	0
uwes3 [a]	1	1	1	1	1	.651	.709	.581	2	0	0
uwes4	1	1	1	1	2	.630	.631	.623	8	0	0
uwes7	1	1	1	1	2	.603	.683	.493	6	0	0
uwes6	1	1	1	0	0	.564	.586	.537	12	0	0
uwes8 [a]	1	1	1	2	0	.549	.582	.508	7	0	0
uwes9	1	0	0	2	0	.377	.454	.305	9	0	0
Total	-	-	-	-	-	.554	.604	.492	-	-	-

Note: AISP: identification of item-derived scales; 0 means that the item does not group into any. Hi: scalability for each item. a Items of the shorter UWES-3 version.

Figure 1.

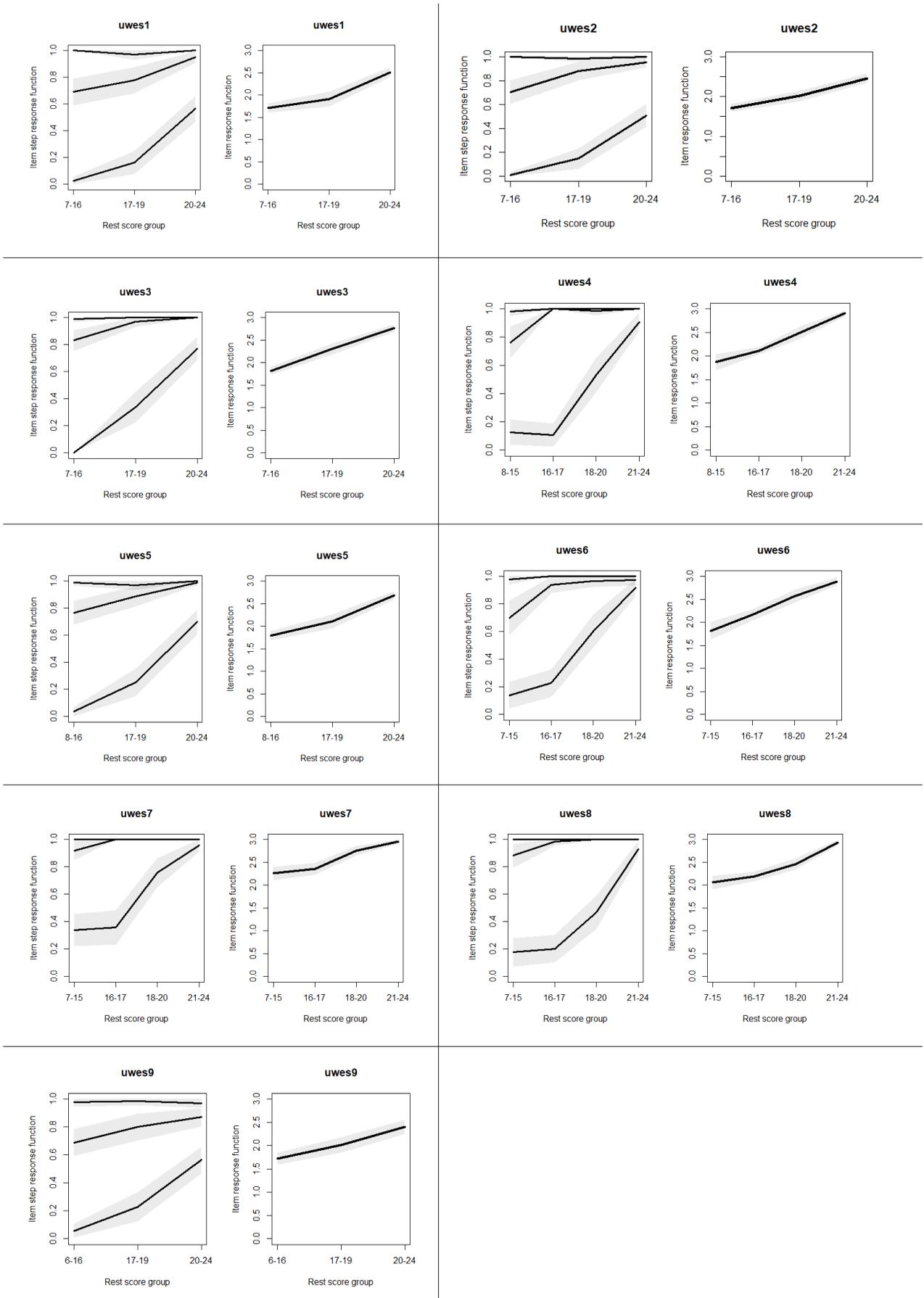*Item Monotonicity: Response Options and Item Score*

Table 3.
SEM Modeling of UWES-9 and UWES-3

| | UWES – 9 | | | | | | UWES – 3 | | | |
| | Tridimensional | | | | Unidimensional | | FT | Thresholds | | |
	VI	DED	AB	R²	FT	R²		τ_1	τ_2	τ_3
uwes1a	.918			.842	.869	.755	.668	-2.143	-.908	.593
uwes2	.880			.775	.844	.712	-	-	-	-
uwes5	.891			.794	.789	.622	-	-	-	-
uwes3 a		.959		.919	.912	.832	1.126	-2.656	-1.497	.297
uwes4		.905		.819	.876	.768	-	-	-	-
uwes7		.812		.660	.783	.614	-	-	-	-
uwes6			.875	.765	.747	.559	-	-	-	-
uwes8 a			.866	.750	.756	.571	.521	-1.916	.05	-
uwes9			.617	.381	.529	.280				
Correlations										
VI	1.0				-	-	-	-	-	-
DED	.717	1.0			-	-	-	-	-	-
AB	.470	.707	1.0		-	-	-	-	-	-
Reliability										
α (IC 95%)					.877 (.84, .89)		.703 (.62, .76)			
ω (IC 95%)					.871 (.83, .89)		.754 (.70, .81)			

Note: VI: vigor. DED: dedication. AB: absorption. TF: total factor. Items of the shorter UWES version. and w: reliability coefficients. a Items of the shorter UWES-3 version.

Reliability

From the MSA model, the MS reliability for the UWES-3 was .892, a level that can be considered high. Likewise, the alpha (α) and omega (ω) reliabilities were moderately high (≥ .70), and slight discrepancy is observed between the two (Table 3, bottom).

Equivalence Between Versions

The unadjusted (r = .931, p < .01) and adjusted (r = .824, p < .01) correlation between UWES-9 and UWES-3 was high, indicating similar ordering of subjects from both scores. Classificatory agreement was moderately high, AC1 = .611 (95% CI = .537, .687), indicating a tendency to differentiate subjects when they are classified into quartiles. Regarding equivalence of construct validity or discrimination with job burnout syndrome, the correlations of the UWES-9 (single score based on the unidimensional model; see paragraphs above) and UWES-3 obtained showed theoretical consistency (Table 4). The correlation with the enthusiasm towards the job dimension of the CESQT-PE was positive and high, while it was negative in direction and low or moderately low with the rest of the scores, as expected. Regarding the correlational difference, both versions produced similar correlational magnitudes; the magnitude of these estimated differences was small, because a) it was between

8% and 11% lower with respect to the UWES-9 version, b) the confidence intervals for the correlational differences (Δr) included zero (except in enthusiasm towards the job), and c) the standardized difference (q) was between .020 and .083, that is to say, of insubstantial magnitude. Considering the age and sex of the participants, the scores of both versions correlated zero in the population, and the correlational difference was insubstantial (q around zero) (Table 4).

Discussion

The aim of the present study was to analyze the psychometric properties of the ultra-brief version of the UWES-3 and to analyze its equivalence with the UWES-9 in a sample of Mexican university professors. The methodological strategy included an exploration of dimensionality and an examination of structural properties through parametric and nonparametric estimations, as well as an equivalence analysis of direct and adjusted correlation patterns, both between the two versions (UWES-3 and UWES-9) and with CESQT dimensions, sex, and age.

The chosen methodological strategy has advantages over previous studies, specifically in the inclusion of adjusted correlation estimates, omitted in the study by Schaufeli et al. (2017) and the estimation of nonparametric results together with the parametric results, omitted in the study by Domínguez-Lara et al. (2021). Regarding the adjusted correlations, controlling the error variance by means of the Levy (1967) correction applied to the correlation between both versions is a practice that is recommended to obtain valid equivalence conclusions (Merino-Soto & Angulo-Ramos, 2013; Smith et al., 2000), due to the fact that spurious correlational sizes tend to be obtained without this correction. On

Table 4.

Correlations Between Versions of the UWES with CESQT

	Correlations		Correlational difference		M	DE
	UWES-9	UWES-3	Δ_r (IC 95%)	q		
Sex	-.082	-.109	(-.018, .072)	.027	-	-
Age	.068	.035	(-.012, .078)	.033	-	-
CESQT						
Enthusiasm towards the job	.618**	.564**	(.017, .095)*	.083	3.33	.653
Indolence	-.245**	-.226**	(-.064, .025)	.020	.67	.555
Exhaustion	-.384**	-.413**	(-.013, .072)	.034	1.75	.976
Guilt	-.188**	-.168**	(-.065, .025)	.020	.62	.620
M	29.66	9.85	-	-	-	-
SD	4.07	1.47	-	-	-	-

Note. Δ_r: *difference of dependent correlations; q: magnitude of the difference; CESQT: dimensions of the Burnout Syndrome Evaluation Questionnaire.* *p < .05; **p < .01

the other hand, the application of nonparametric modeling is generally considered as a precursor of parametric models, because it does not require a known distribution of the construct (e.g., normal distribution) (Mokken, 1971; Van der Ark, 2012).

The dimensional and structural parametric and nonparametric analysis coincided in the confirmation of a one-dimensional structure of the UWES-9, where only the three items of the brief version presented the highest factor loadings and a completely satisfactory parametric fit. This confirms the essential one-dimensionality of the UWES-3, with high construct validity of the items. On the other hand, internal consistency was acceptable for both versions, although higher for the UWES-9. The decrease in the reliability of the UWES-3 scores is a consequence of the reduced number of items and its consequent restriction of construct variance. This is not necessarily a problem in a relative view, because the appropriate or reasonable magnitude of a score's reliability is conditioned using this score and the context of application (American Educational Research Association [AERA], American Psychological Association [APA], & National Council for Measurement in Education [NCME], 2014). Therefore, with a reliability magnitude around .75 of the UWES-3 scores, it seems to suit screening assessments, monitoring, psychosocial surveillance, and group descriptions. For additional uses requiring greater scoring accuracy, the UWES-9 version may be better suited. As a side note, the difference between the alpha and omega coefficient of the UWES-3 is a consequence of the distance of the factor loading of item 3 from the other two items. One implication is that, if this remains the same in other studies, then the appropriate coefficient for estimating reliability is ω, not α.

One of the findings of the study was the magnitude of the scaling coefficients H in the female group compared to the male group. The scaling coefficients H are interpretable as coefficients of discrimination of an item with respect to the mean score (Van Abswoude et al., 2004) and partially refer also to the item measurement quality. This implies that the scaling and overall quality properties of the UWES-3 in

females are not as strong as in the case of males. Results such as those obtained here, where group differences are apparent in measurement quality, require further research (Wind, 2017). In item differential functioning, within the Mokken Scale Analysis (MSA) framework, the intersection of curves characteristic of the response options may be behind these differences, demonstrating lower ability to discriminate individual differences in females. It is likely that these differences are idiosyncratic to the sample of participants and therefore not replicable; however, this alerts us to the need to include other indicators in a larger sample size condition.

Regarding the equivalence of the UWES-3 and UWES-9 versions, the correlations between both versions presented coefficients higher than r = .82, with a moderately high classificatory agreement, and the patterns of correlations with the CESQT and demographic variables (sex, age) were practically similar both theoretically and empirically, in the same direction and with statistically insignificant differences in size. This leads to the conclusion that the UWES-3 is psychometrically equivalent to the UWES-9 for the assessment of work engagement in university professors.

It is relevant to point out that the ultrashort 3 item version (UWES-3) represents in each item the same three important conceptual aspects of engagement that were coined from its origin: vigor, dedication, and absorption, so content validity is not compromised, nor is there a loss of theoretical representation built on its roots. In general, there is no significant loss of information or variance with respect to versions of the UWES with more items, which implies implicit practical benefits.

In this regard, there is currently a trend to investigate the feasibility of ultra brief measures due to the practical benefits that this has; the advantages are multiple and include, among others, taking less time of participants and institutions in filling out questionnaires, decreasing attrition rates due to long questionnaires, avoiding respondent fatigue or monotony (which at the same time ensures the validity of the measurements), and not least, facilitating their inclusion in large epidemiological studies or national surveys

measuring the same constructs without many items. These benefits can be extended to the short version of the UWES-3 analyzed in this study an option of a more practical and simple assessment of work engagement in university professors, although more studies seem necessary to confirm the trend found so far.

Limitations of the study could be identified in the non-random sample size and strategy, the choice of few variables for discrimination (CESQT dimensions, age, sex) and the implicit lack of independence in the use of UWES-9 and UWES-3 in the same study, making it difficult to understand their true reliability and validity in a truly independent manner (Schaufeli et al., 2017). However, the nonparametric and robust estimates used may mitigate the sampling problems to some extent. Also, the choice of burnout—measured using the CESQT—as a criterion for equivalence analysis may have been sufficient given the expected theoretical divergence between the two constructs, considering that engagement was born and gained identity as just an opposite of burnout. It is up to future studies to analyze the pattern of correlations with other variables in other samples using jointly and independently the UWES-9 and the UWES-3 to confirm such equivalence in different scenarios. In the meantime, this study serves as an initial contribution in this sense for the benefit of the measurement of engagement in the Latin-American teaching workforce.

References

Álvarez Garón, S. M., & Peña Fuentes, J. N. (2019). Adaptación del Utrecht Work

Engagement Scale–UWES en profesores universitarios de Villavicencio (Tesis de pregrado) *Universidad Cooperativa de Colombia, Villavicencio*. http://hdl.handle.net/20.500.12494/12393

American Educational Research Association, American Psychological Association & National Council for Measurement in Education [AERA, APA & NCME] (2014). The Standards for Educational and Psychological Testing. Washington, D.C.: AERA

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods, 24*(4), 468–491. https://doi.org/10.1037/met0000200

Burić, I., & Macuka, I. (2018). Self-efficacy, emotions and work engagement among teachers: A two wave cross-lagged analysis. *Journal of Happiness Studies*, 19(7), 1917-1933. https://doi.org/10.1007/s10902-017-9903-9

Calderón-de la Cruz, G. A., Merino-Soto, C., Juárez-García, A., Dominguez-Lara, S., & Fernández-Arata, M. (2020). ¿Es replicable la estructura factorial del Maslach Burnout Inventory Human Service Survey (MBI-HSS) en la profesión de enfermera del Perú?: un estudio nacional. *Enfermería Clínica*, 30(5), 340-348. https://doi.org/10.1016/j.enfcli.2019.12.013

DeCarlo, L. T. (*1997*). On the meaning and use of kurtosis. *Psychological Methods, 2*(3), 292-307. https://doi.org/10.1037/1082-989X.2.3.292

Diedenhofen, B. & Musch, J. (2015). A Comprehensive Solution for the Statistical Comparison of Correlations. *PLoSONE, 10*(4), e0121945. http://doi.org/10.1371/journal.pone.0121945

Dominguez-Lara, S. A., Fernández-Arata, M., & Seperak-Viera, R. (2021). Análisis psicométrico de una medida ultra-breve para el engagement académico: UWES-3S: UWES-3S en estudiantes universitarios. *Revista Argentina de Ciencias del Comportamiento*, 13(1), 25-37.https://doi.org/10.32348/1852.4206.v13.n1.27780

Flores Jiménez, C., Fernández Arata, M., Juárez García, A., Merino Soto, C., & Guimet Castro, M. (2015). Entusiasmo por el trabajo (engagement): un estudio de validez en profesionales de la docencia en Lima, Perú. *Liberabit*, 21(2), 195-206. http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1729-48272015000200003

Gil-Monte, P. R. (2011). Evaluación del cuestionario CESQT. Madrid.

Gil-Monte, P. R., & Noyola, C. V. (2011). Estructura factorial del "Cuestionario para la Evaluación del Síndrome de Quemarse por el Trabajo" en maestros mexicanos de educación primaria. *Revista Mexicana de Psicología*, 28(1), 75–84. http://www.redalyc.org/articulo.oa?id=243029630007

Gil-Monte, P. R., Unda, S. R., & Sandoval, O. J. I. (2009). Validez factorial del « Cuestionario para la Evaluación

del Síndrome de Quemarse por el Trabajo » (CESQT) en una muestra de maestros mexicanos. *Salud Mental*, 32(3), 205–214. http://www.revistasaludmental.mx/index.php/salud_mental/article/view/1285

González-Romá, V., Schaufeli, W.B., Bakker, A. & Lloret, S. (2006). Burnout and engagement: Independent factors or opposite poles? *Journal of Vocational Behavior, 68,* 165-174. https://doi.org/10.1016/j.jvb.2005.01.003

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, 61, 29-48. https://doi.org/10.1348/000711006X126600

Gwet, K. L. (2019). irrCAC: Computing Chance-Corrected Agreement Coefficients (CAC). R package version 1.0. https://CRAN.R-project.org/package=irrCAC

Han, J., Yin, H., Wang, J., & Zhang, J. (2020). Job demands and resources as antecedents of university teachers' exhaustion, engagement and job satisfaction. *Educational Psychology*, 40(3), 318-335. https://doi.org/10.1080/01443410.2019.1674249

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2020). semTools: Useful tools for structural equation modeling. R package version 0.5-3. https://CRAN.R-project.org/package=semTools

Juárez-García, A. (2015). Investigaciones psicométricas de escalas psicosociales en trabajadores mexicanos. México, DF: Plaza y Valdés.

Levy, P. (1967) The correction for spurious correlation in the evaluation of short-form tests. *Journal of Clinical Psychology, 23*, 84-86. https://doi.org/10.1002/1097-4679(196701)23:1<84::aid-jclp2270230123>3.0.co;2-2

Loevinger, J. (1948). The technic of homogeneous tests compared with some aspects of scale analysis and factor analysis. Psychological Bulletin, 45(6), 507-529. https://doi.org/10.1037/h0055827

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science of India, 12*(1), 49-55. http://hdl.handle.net/10263/6765

Makowski, (2018). The *psycho* Package: an efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software, 3*(22), 470. https://doi.org/10.21105/joss.0047

Mangiafico, S. (2021). rcompanion: Functions to support extension education program evaluation. R package version 2.4.1. https://CRAN.R-project.org/package=rcompanion

Martins, E. L. D. S., & Mendonça, H. (2019). Evidencias de validez de la escala de compromiso de docentes de educación básica. *Psicologia em Pesquisa*, 13(3), 159-172. http://dx.doi.org/10.34019/1982-1247.2019.v13.27110

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

Maslach, C., Jackson, S.E., Leiter, M.P., Schaufeli, W.B. and Schwab, R.L. (1986) Maslach burnout inventory instruments and scoring guides forms: General, human services, & educators. *Health and Quality of life Outcomes*, 7, 31. http://www.mindgarden.com

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

Merino-Soto, C., & Angulo-Ramos, M. (2013). Validación en Chile de la escala de sobrecarga del cuidador de Zarit en sus versiones original y abreviada: corrección. *Revista Médica de Chile, 141*, 1083-1084. http://dx.doi.org/10.4067/S0034-98872013000800019

Mokken, R. J. (1971) *A Theory and Procedure of Scale Analysis*. Berlin, Germany: De Gruyter.

Molenaar, I. W., & Sijtsma, K. (1984). Internal consistency and reliability in Mokken's nonparametric item response model. *Tijdschrift voor onderwijsresearch, 9*, 257–268.

Molenaar, I. W., & Sijtsma, K. (1988). Mokken's approach to reliability estimation extended to multicategory items. *Kwantitatieve methoden, 9*(28), 115-126. https://research.tilburguniversity.edu/en/publications/mokkens-approach-to-reliability-estimation-extended-to-multicateg

Revelle, W. (2020) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois. R package version 2.0.12. https://CRAN.R-project.org/package=psych

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. URL http://www.jstatsoft.org/v48/i02/.

Ruscio, J. (2018). RGenData: Generates Multivariate Nonnormal Data and Determines How Many Factors to Retain. R package version 1.0. https://CRAN.R-project.org/package=RGenData

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. Psychological Assessment, 24(2), 282–292.

https://doi.org/10.1037/a0025697

Salanova, M., & Schaufeli, W. (2009). El engagement en el trabajo: cuando el trabajo se convierte en pasión. Alianza Editorial; Madrid.

Salanova, M., Schaufeli, W.B., Llorens, S., Peiro, J.M. & Grau, R. (2000). Desde el 'burnout' al 'engagement': ¿ Una nueva perspectiva? [From 'burnout' to 'engagement': A new perspective?] *Revista de Psicología del Trabajo y las Organizaciones, 16*, 117-134. https://journals.copmadrid.org/jwop/art/7c590f-01490190db0ed02a5070e20f01

Schaufeli, W. B., & Bakker, A. B. (2010). Defining and measuring work engagement: Bringing clarity to the concept. In A. B. Bakker & M. P. Leiter (Eds.), Work engagement: A handbook of essential theory and research (pp.10-24). New York: Psychology Press

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

Schaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness studies*, 3(1), 71-92.

Schaufeli, W. B., Shimazu, A., Hakanen, J., Salanova, M., & De Witte, H. (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. (2017). General engagement: Its conceptualization and measurement. *Journal of Well-Being Assessment, 1, 9-24.*

Schaufeli, W.B. (2018). *Burnout in Europe: Relations with national economy, governance, and culture.* Research Unit Occupational & Organizational Psychology and Professional Learning (internal report). KU Leuven: Belgium.

Sijtsma, K., & Van der Ark, L. A. (2017). A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. *British Journal of Mathematical and Statistical Psychology, 70*(1), 137-158. https://doi.org/10.1111/bmsp.12078

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12*(1), 102–111. https://doi.org/10.1037/1040-3590.12.1.102

Steiger, J. H. (1980). Testing pattern hypotheses on correlation matrices: alternative statistics and some empirical results. *Multivariate Behavioral Research, 15*, 335-352. https://doi.org/10.1207/s15327906mbr1503_7

Straat J. H., Van der Ark L. A., Sijtsma K. (2016). Using conditional association to identify locally independent item sets. *Methodology* 12 117–123. https://doi.org/10.1027/1614-2241/a000115

Straat, J. H., Van der Ark, L. A., & Sijtsma, K. (2013). Comparing optimization algorithms for item selection in Mokken scale analysis. *Journal of Classification, 30,* 72-99. https://doi.org/10.1007/s00357-013-9122-y

Straat, J. H., Van der Ark, L. A., and Sijtsma, K. (2013). Comparing optimization algorithms for item selection in Mokken scale analysis. *Journal of Classification, 30,* 72-99. https://doi.org/10.1007/s00357-013-9122-y

Van Abswoude, A.A.H., Van der Ark, L.A., & Sijtsma, K. (2004). A comparative study of test dimensionality assessment procedures under nonparametric IRT models. Applied Psychological Measurement, 28, 3–4. http://dx.doi.org/10.1177/0146621603259277

Van der Ark, L. A. (2012). New developments in Mokken Scale Analysis in R. Journal of Statistical Software,48(5), 1-27. http://www.jstatsoft.org/v48/i05/

Van Schuur, W.H. (2003). Mokken Scale Analysis: between the Guttman scale and parametric item response theory. *Political Analysis*, *11*, 139-163. https://doi.org/10.1093/pan/mpg002

Wetzel, E., Böhnke, J. R., & Brown, A. (2016). Response biases. In F. T. L. Leong, D. Bartram, F. M. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international handbook of testing and assessment* (p. 349–363). Oxford University Press.

Wind, S.A. (2017). An Instructional Module on Mokken Scale Analysis. Educational Measurement: Issues and Practice, 36: 50-66. https://doi.org/10.1111/emip.12153

World Medical Asociation (AMM). (2013). Declaración de Helsinki de la AMM - Principios éticos para las investigaciones médicas en seres humanos. *World Medical Association, Inc*. https://doi.org/10.1001/jama.2013.281053

Wyatt, M., & Dikilitaş, K. (2016). English language teachers becoming more efficacious through research engagement at their Turkish university. *Educational*

Action Research, 24(4), 550-570. http://dx.doi.org/10.1080/09650792.2015.1076731

Zhu, N. Q. (2001). *The effects of teachers' flow experiences on the cognitive engagement of students.* (Tesis Doctoral). Universidad de San Diego.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, *12*, 399-413. http://doi.org/10.1037/1082-989X.12.4.399