

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

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

Citar como: Juárez-García, A.; Merino-Soto, C.; Villamar Sánchez, D. (2023). Más breve puede ser mejor: Propiedades métricas de la UWES – 3 ítems para evaluar “engagement” en profesores universitarios mexicanos. *Revista Digital de Investigación en Docencia Universitaria*, 17(2), e1623. <https://doi.org/10.19083/ridu.2023.1623>

Recibido: 21/04/2022. **Revisado:** 29/04/2022 **Aceptado:** 02/02/2023. **Publicado:** 30/06-2023.

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 educación (CESQT-PE) con el propósito de estimar su validez discriminante. **Resultados:** Los resultados confirmaron correlaciones muy altas entre ambas versiones 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 ($\alpha = .02-.08$). **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

Shorter maybe better: Psychometric properties of the UWES-3 items for measuring engagement in university Mexican teachers

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

*Correspondencia:

Arturo Juárez García
arturojuarezg@hotmail.com

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.

Introducción

El *engagement* laboral es quizás el efecto psicológico laboral con enfoque positivo más popular en últimas las décadas. Aunque existen intentos de su traducción al castellano en conceptos tales como “entusiasmo laboral” (Juárez-García, 2015), su acepción en inglés sigue siendo la más común, tanto en ámbitos de consultoría como en contextos académicos, quizás para no confundirse con otros fenómenos psicológicos. Maslach y Leiter (1997) lo posicionaron justo como un opuesto directo al síndrome de *burnout*: esto es, sentirse energético en contraparte al síntoma de agotamiento emocional, involucrarse psicológicamente en la actividad laboral en contraparte a la despersonalización del *burnout*; y finalmente, sentirse auto eficaz profesionalmente en oposición a la ausencia de tal certeza en las capacidades cuando hay desgaste psicológico. Sin embargo, este paradigma padece de un enfoque restringido a dos polos opuestos mutuamente excluyentes, pues su lógica básica es que bajos niveles de *burnout* representan en automático altos niveles de *engagement* y por ende ambos pueden medirse con el mismo instrumento (Maslach Burnout Inventory -MBI-) (Maslach, Jackson, Leiter, Schaufeli & Schwab, 1986), lo que evidentemente ha recibido críticas (González-Roma, Schaufeli, Bakker & Lloret, 2006).

De acuerdo con Schaufeli, Salanova, González-Roma y Bakker (2002), el *engagement* laboral

es un fenómeno correlacionado negativamente al *burnout*, pero al mismo tiempo independiente. Lo definen como un efecto motivacional positivo que se da por el trabajo y se caracteriza por tres componentes: 1) Vigor, que se caracteriza por los altos niveles de energía y resiliencia mental mientras se trabaja, voluntad de esforzarse y persistir en situaciones difíciles en el trabajo; 2) Dedicación, que se refiere a estar fuertemente motivado en el trabajo y experimentar un sentido de significancia, inspiración, orgullo y desafío; y 3) Absorción, que se refiere al hecho estar completamente concentrado e inmerso en el trabajo, con la sensación de que el tiempo “pasa volando”, en una especie de vinculación plena con el trabajo (Salanova & Schaufeli, 2009). El *engagement* ha demostrado su relación con la motivación, la satisfacción, la felicidad, la iniciativa y el buen desempeño laboral entre diversos efectos individuales y organizacionales favorables en diversos contextos (Salanova & Schaufeli, 2009). Incluso, se han encontrado asociaciones del *engagement* con la eficacia productiva de los países, el producto interno bruto, la presencia de integridad y ausencia de corrupción, así como con culturas con alto individualismo, alto disfrute de la vida y baja distancia de poder (Schaufeli, 2018).

El *engagement* laboral en la tarea docente ha sido poco investigado, sin embargo, se han demostrado ya las relaciones positivas de este fenómeno con la satisfacción de los profesores (Han, Yin, Wang, & Zhang, 2020), la presencia de más emociones positivas mientras trabajan (Burić & Macuka, 2018), mayor eficacia en su actividad laboral (Wyatt & Dikilitaş, 2016), e incluso, con el mejoramiento de desempeño cognitivo de sus estudiantes (Zhu, 2001), entre muchas otras consecuencias positivas.

Un parteaguas en el desarrollo del constructo se instaló en los trabajos de Schaufeli y colaboradores (Schaufeli et al., 2002), quienes no solo demostraron que el constructo era independiente al *burnout* (Gonzalez-Roma et al., 2006), sino que lo definieron y diferenciaron conceptualmente de otros constructos con base en evidencia empírica. Además, propusieron una medida con satisfactorias propiedades psicométricas en distintas muestras y regiones: el *Utrecht Engagement Scale*

(UWES) ([Salanova, Schaufeli, Llorens, Peiro & Grau, 2000](#); [Schaufeli et al., 2002](#); [Schaufeli, 2017](#)).

La escala UWES ha sido validada en diversos idiomas en más de 30 países en los cinco continentes ([Schaufeli & Bakker, 2010](#)). Su versión inicial de 24 ítems se validó en una propuesta de 17 ítems funcionales, con la estructura trifactorial esperada e índices de consistencia interna (α) superiores a .70 en poblaciones de estudiantes y trabajadores ([Schaufeli et al., 2002](#)). Posteriormente [Salanova et al., \(2000\)](#), probaron una versión de la UWES de 15 ítems en una muestra de trabajadores que utilizan las tecnologías de la información, a la que dedicaron cinco preguntas para vigor, cinco para dedicación y cinco para absorción en una escala de siete opciones de respuesta de frecuencia Likert que va desde "0=Nunca", hasta "6=Todos los días". Los resultados confirmaron la estructura tridimensional e índices de consistencia interna (α) superiores a .73 en todas las subescalas, así como la independencia con el constructo de burnout y relaciones estadísticamente significativas en las direcciones esperadas con indicadores organizacionales tales como el compromiso organizacional, la satisfacción o el gozo laboral.

En una siguiente etapa, [Schaufeli, Bakker y Salanova \(2006\)](#) validaron una versión de la UWES de 9 ítems (3 para cada componente) en 10 países europeos de diferentes grupos ocupacionales, la consistencia interna fue satisfactoria ($\alpha > .70$) y aunque la estructura trifactorial se comprobó, los mejores índices fueron para una estructura unidimensional. Dadas las altas correlaciones entre los tres componentes, los autores sugirieron el uso del puntaje total en una medida única global de *engagement*.

Más recientemente, [Schaufeli, Shimazu, Hakanen, Salanova & De Witte \(2019\)](#), validaron una versión ultra-breve de solo tres ítems (UWES-3) en muestras de varios países, la cual mantuvo buenos índices psicométricos, compartió hasta un 92% de varianza con la versión de nueve ítems (UWES-9) y no presentó diferencias en los patrones de correlación con otras variables. Los autores sugieren el uso del UWES-3 en estudios futuros e incorporarlo en encuestas nacionales, por todas las ventajas prácticas que representa.

Aunque en Latinoamérica existen algunas validaciones de la escala de *engagement* laboral

(UWES) en muestras de docentes ([Flores, Fernández, Juárez, Merino y Guimet, 2015](#); [Álvarez Garzón & Peña Fuentes, 2019](#); [Martins & Mendonça, 2019](#)), estos se han realizado con las versiones de 17, 15 ó 9 ítems y solo se identificó una publicación muy reciente que validó la versión ultra-breve UWES-3 en estudiantes peruanos. Se trata del estudio peruano de [Dominguez, Fernández y Seperak \(2021\)](#) quienes concluyeron que las propiedades psicométricas de esta versión breve no diferían a las otras versiones del UWES. No obstante lo anterior, siguen pendientes estudios de validación de esta versión ultra-breve en poblaciones de trabajadores académicos en poblaciones latinas, en particular en México, donde no hay estudios al respecto. Las implicaciones prácticas y de investigación de contar con una medida ultra-breve para el tamizaje del fenómeno de *engagement* en docentes, representa múltiples ventajas asociadas a una detección sencilla, rápida y con una relación lo más positiva posible en términos de efectividad/costo y de eficiencia/tiempo en la evaluación de este fenómeno en profesores.

Dado lo anterior, el objetivo del presente estudio es analizar las propiedades psicométricas de la versión ultra-breve UWES-3 y analizar su equivalencia con la UWES-9 en una muestra de docentes universitarios mexicanos.

Método

Participantes

Considerando una población total de 1,825 profesores e investigadores en los centros temáticos y regionales de la Universidad de Guadalajara, se envió un correo electrónico a todos ellos con la invitación a participar y un vínculo de acceso a la encuesta a través de la plataforma SurveyMonkey®. Para ingresar a contestar el cuestionario fue necesaria la aceptación de la carta de consentimiento. Aceptaron participar 259 académicos, por lo que la tasa de respuesta fue del 14.19%. Los criterios de inclusión fueron: realizar actividades de investigación y docencia, ser un trabajador activo en la Universidad al momento del estudio, tener un contrato de tiempo com-

pleto, aceptar participar en el estudio y firmar la carta de consentimiento informado. Los criterios de exclusión fueron: ser docentes en sabático, ser profesores con contrato distinto al tiempo completo y no participar en algún programa estímulo. El criterio de eliminación fue no contestar la batería de instrumentos o contestarla parcialmente. Finalmente, conforme a estos criterios se obtuvo una muestra de 247 académicos.

La muestra final estuvo distribuida de la siguiente manera: por 122 hombres (49.4%) y 125 mujeres (50.6%), con una edad promedio de 49.28 años (Desviación estándar = 9.677). Respecto al estado civil, 149 fueron casados (60.3%), 53 solteros (21.5%), 23 viven en unión libre (9.3%), 18 están divorciados (7.3%) y 4 son viudos (1.6%). En nivel máximo de estudios se observa que 38 tienen maestría (15.4%), 179 tienen doctorado (72.5%), y 30 tienen postdoctorado (12.1%). Respecto al centro universitario de adscripción 205 pertenecen a un Centro Temático (82.9%) y 42 a un Centro Regional (17.1%).

Conforme al nombramiento oficial, 240 participantes señalaron ser Profesor-Investigador (97.2%) y 7 señalaron ser Técnico Académico (2.8%). Sin embargo, todos se dedican a realizar actividades de investigación y docencia. El promedio de antigüedad en la institución fue de 19.32 años (Desviación estándar= 10.983), aunque el promedio en el puesto actual fue de 8.81 años (Desviación estándar = 7.832).

Consideraciones éticas

Se formuló una carta de consentimiento informado. A través de este documento se solicitó la participación voluntaria de los académicos, se garantizó la confidencialidad de su identidad y se les confirió la libertad de abandono de la encuesta en el momento deseado. No hubo identificadores personales, ni algún procedimiento invasivo. Se siguieron todas las recomendaciones de la declaración de Helsinki ([World Medical Association \[AMM\], 2013](#))

Instrumentos

En la misma lógica de los estudios disponibles sobre la validez y equivalencia de la versión corta UWES-3, se seleccionó como criterio concurrente la versión de 9 ítems por ser la versión mejorada de

versiones previas y la que posee mejores capacidades psicométricas que la versión de 17 ó 15 ítems (Schaufeli et al., 2006). Asimismo, como criterio externo para valorar los patrones de equivalencia de correlaciones entre ambas versiones con un criterio externo, se consideró la edad y el sexo, así como una medida de *burnout* o desgaste emocional con la que se espera un tipo de validez o correlación divergente con las medidas del UWES, como sucedió en estudios iniciales. En este estudio se utilizó el CESQT, un instrumento que denomina síndrome de quemarse por el trabajo al *burnout*. Tales instrumentos se desglosan a continuación:

Utrecht Work Engagement Scale (UWES) en sus versiones de 9 y 3 ítems ([Schaufeli et al., 2017](#)).

Es una medida de evaluación del *engagement*, dentro del mismo modelo en que se crearon sus versiones anteriores (i.e., UWES-15 y UWES-9). La versión ultra-breve UWES-3 está compuesta por tres ítems que representan a cada uno de los componentes originales, referidos a sentir energía (ítem 1, Vigor), entusiasmo (ítem 2, Dedicación) y e inmersión (ítem 8, Absorción). Sus opciones de respuesta están escaladas ordinalmente de 0 (*Nunca-Ninguna vez*) a 6 (*Siempre-Todos los días*). La confiabilidad de consistencia interna (α) en el estudio original osciló entre un de .77 y .85.

Cuestionario de Evaluación de Síndrome de Quemarse por el Trabajo para personal de educación (CEQST-PE). Contiene 20 ítems distribuidos en 4 dimensiones ([Gil-Monte, 2011](#)). La Ilusión por el trabajo (1) implica el deseo del individuo de alcanzar las metas laborales porque suponen una fuente de satisfacción personal. El Desgaste psíquico (2), se refiere al Agotamiento emocional y físico debido a que en el trabajo se tiene que tratar a diario con personas que presentan o causan problemas. La Indolencia (3) representa la presencia de actitudes negativas de indiferencia y cinismo hacia los clientes de la organización (pacientes, alumnos, etc.). La Culpa (4), involucra sentimientos de culpa que la persona tiene por el comportamiento y las actitudes negativas desarrolladas en el trabajo, en especial hacia las personas con las que se establecen relaciones laborales. Las opciones de respuesta incluyen una escala de frecuencia de cinco grados que van de 0 (nunca) a 4 (Muy frecuentemente; todos los días) ([Gil-Mon-](#)

te & Noyola, 2011). El CESQT-PE ha mostrado validez y confiabilidad satisfactoria en muestras mexicanas y latinas (Gil-Monte, Unda & Sandoval, 2009) y con eficacia superior al *Maslach Burnout Inventory - Human Services Survey* (MBI-HSS) (Juárez-García, 2015, Calderón-de la Cruz, Merino-Soto, Juárez-García, Dominguez-Lara & Fernández-Arata, 2020), por lo que la escala y sus dimensiones, son un criterio idóneo en la validez discriminante del UWES-3.

Procedimiento

Análisis. Consistió en dos procedimientos, el primero preliminar para prevenir el efecto de distorsiones no intencionadas de respuesta (Wetzel, Böhnke, & Brown, 2016) y para describir las respuestas a los ítems, y el segundo enfocado en el objetivo principal del estudio.

Análisis preliminar. En el primero, se escanearon los datos para detectar posibles sesgos de respuestas expresados por valores atípicos multivariados (Meade, & Craig, 2012). Considerando su eficiencia (Meade, & Craig, 2012), se utilizó la distancia D_2 (Mahalanobis, 1936) con el programa *normtest* (DeCarlo, 1997). Los participantes detectados como valores atípicos fueron removidos de la base de datos. Luego en el nivel de los ítems se obtuvieron estadísticos descriptivos y distribucionales, y correlacionales (igualdad de las correlaciones inter-ítem; Steiger 1980), en los programas R *psych* (Revelle, 2020), y *rcompanion* (Mangiafico, 2021)

Análisis psicométrico. Para el objetivo principal, el análisis de las propiedades psicométricas fue secuencialmente realizado mediante la evaluación de la dimensionalidad (i.e., número de factores latentes), la escalabilidad de los puntajes e ítems observados, las características estructurales, la equivalencia entre grupos, y la correlación con otras variables.

Dimensionalidad. Debido que existe discrepancia entre la representación de constructo más apropiada con el UWES, primero se exploró el número de variables latentes mediante la evaluación del número de factores latentes. Se implementó un procedimiento de consenso entre 9 métodos de extracción de factores disponibles (e.g., análisis paralelo, coordinadas óptimas,

MAP-Velicer, etc.), mediante el programa *psycho* (Makowski, 2018). También se utilizó el método de *Comparación de Datos* (CD; Ruscio, & Roche, 2012; programa *RGenData*, Ruscio, 2018) debido que tiende a mostrar adecuada precisión (Auerwald, & Moshagen, 2019).

Propiedades estructurales. Se aplicaron dos enfoques, uno no paramétrico apropiado para medidas con pocos ítems (Van Schuur, 2003), y otro paramétrico, analizados en ese orden. Respecto al modelo no paramétrico, se corroboró la dimensionalidad predominante hallada en el análisis anterior mediante el *Mokken Scale Analysis* (MSA; Mokken, 1971). Como un proxy del número de dimensiones latentes (Straat, Van der Ark, & Sijtsma, 2013), se aplicó el algoritmo Automated Item Selection Procedure (AISP; Straat, Van der Ark, & Sijtsma, 2013) para seleccionar ítems escalables dentro de grupos homogéneos, e identificar así el número de escalas. Se utilizó el rango de .30 hasta .70, en pasos de .10 unidades del coeficiente de escalabilidad H para la detección de grupos homogéneos de ítems. Una vez completado esto, en el siguiente paso se corroboraron las propiedades de los puntajes observados, que son precursoras para el modelamiento paramétrico más complejo (e.g., basados en SEM o IRT). De este modo, se verificó el cumplimiento del *modelo de homogeneidad monótonica* (MHM; Sijtsma, & Van der Ark, 2017), mediante la evaluación de: a) la escalabilidad con el coeficiente H (Loevinger, 1948) para el puntaje y los ítems (Hi); coeficientes mayores a .40 son aceptables; Sijtsma, & Van der Ark, 2017). Luego, b) la independencia local, mediante los coeficientes $W^{(1)}$ y $W^{(2)}$ para las relaciones inter-ítem, y $W^{(2)}$ para cada ítem (Straat, Van der Ark, & Sijtsma, 2016); finalmente, c) la monotonidad de los ítems (i.e., la relación monótonica incremental entre los ítems y el puntaje de la escala), evaluado mediante gráficos de curvas características de los ítems. Finalmente, desde el MSA la confiabilidad se estimó con el coeficiente MS. Para el análisis MSA, se usó el programa *mokken* (Van der Ark, 2012).

En el segundo enfoque, mediante el modelamiento paramétrico lineal, se utilizó el modelamiento de ecuaciones estructurales (SEM), con el estimador WLSMV para variables categóricas,

e índices prácticos de ajuste, como CFI (≥ 95), TLI (≥ 95) y SRMR ($\leq .05$). Se evaluaron tres modelos: dimensiones correlacionadas y unidimensional en el UWES-9, y unidimensional en el UWES-3. Se utilizó el programa lavaan (Rosseel, 2012) y sem-Tools (Jorgensen, Pornprasertmanit, Schoemann, & Rosseel, 2020).

Confiabilidad. Se examinó la confiabilidad de consistencia interna, mediante el coeficiente MS (Molenaar & Sijtsma, 1984, 1988) desde el modelamiento Mokken Scaling Analysis; desde el modelamiento lineal, se estimaron los coeficientes alfa (α) y omega (ω), con intervalos de confianza en 95%.

Equivalencia. La equivalencia entre las versiones (UWES-9 y UWES-3) fue evaluada mediante: a) la correlación ajustada entre sus puntajes directos (Levy, 1967); y el coeficiente AC1 (Gwet, 2008, 2019), para el acuerdo en la clasificación de los participantes en cuartiles. Finalmente, la equivalencia de la validez de constructo se hizo mediante la comparación de correlaciones dependientes (Zou, 2007), entre el UWES-9 y UWES-3 con las escalas de CES-QT y variables demográficas (sexo y edad). Se usó el programa cocor (Diedenhofen & Musch, 2015).

Resultados

Análisis preliminar.

Información descriptiva de los ítems. La tendencia de respuesta del UWES (Tabla 1) se ubicó entre las opciones 3 (“regularmente: algunas veces al mes”) y 4 (“frecuentemente: una vez por semana”), mientras que la primera opción (“esporádicamente: pocas veces al año o menos”) fue predominantemente escasa y hubo una pobre frecuencia de respuesta; esto fue más claro en los ítems 7 y 8, donde no hubo respuestas en la primera opción. Las diferencias en el ranking promedio de cada ítem fueron estadísticamente significativas (Friedman - $\chi^2 = 278$, $gl = 8$, $p < .001$), con tamaño de efecto Kendall - $W = .523$ (IC 95% = .514, .262). Respecto a las correlaciones inter-ítem (no mostrada en la Tabla 1), la identidad correlacional no fue mantenida, $\chi^2 = 2601.53$, $gl = 36$ ($p < .01$), indicando que existen diferencias en la magnitud entre las correlaciones inter-ítem.

Detección de sesgo de respuesta. Mediante el procedimiento Mahalanobis - $D2$, el valor crítico establecidos en la distribución $F(9, 249)$ fue $D2 > 30.67$

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

(nivel .05 con ajuste Bonferroni); de este modo, se removieron de la base de datos 6 participantes con valores $D2$ entre 31.41 y 52.97. El examen visual del patrón de respuesta de estos sujetos mostró inconsistencia de respuesta en todos los ítems del UWES.

Dimensionalidad

La dimensionalidad del UWES-9 en un solo factor fue respaldada por cinco métodos (55.5%; óptimas coordenadas, factor de aceleración, análisis paralelo, MAP-Velicer y VSS complejidad 1), mientras que tres métodos (33.3%) respaldaron dos factores, y solo un método (BIC ajustado por tamaño muestral) indicó tres factores. El método de CD señaló dos posibles factores. La mayor convergencia ocurrió para la unidimensionalidad con los métodos más recomendados en la práctica de investigación (Auerswald, & Moshagen, 2019), y por lo tanto el análisis factorial se orientó hacia la estimación de los parámetros de un modelo unidimensional general de la versión de 9 ítems. Desde otro enfoque, con el enfoque no paramétrico MSA (ver siguiente sección), se procedió a realizar la verificación de la unidimensionalidad.

Mokken Scale Analysis (MSA)

Identificación de escalas. En primer lugar, el procedimiento AISP arrojó con más claridad que se puede obtener una sola escala en el rango de H entre

.30 a .50 (Tabla 2, lado izquierdo). En el nivel más alto ($H > .50$) no se identificó una configuración de tres escalas, lo que converge con el análisis paramétrico del párrafo anterior, en que los ítems convergen en una sola dimensión del UWES-9.

Escalabilidad. El coeficiente H para el puntaje total (Tabla 2, parte central) fue alto ($H = .554$, e.e. = .033, IC 95%: .489, .618); en varones la escalabilidad fue moderadamente mayor ($H = .604$, e.e. = .044, IC 95% = .517, .690) comparado con las mujeres ($H = .492$, e.e. = .048, IC 95% = .397, .586). La escalabilidad inter-ítem (H_{ij}) varió entre .869 (e.e. = .059) y .029 (e.e. = .098), con el ítem 9 mostrando los coeficientes más bajos con el resto de los ítems. Finalmente, la escalabilidad para cada ítem (H_i) fue superior a .540, excepto para el ítem 9. Todos los coeficientes H fueron estadísticamente significativos ($z > 2.50$).

Dependencia local. $W^{(2)}$ varió entre 27.82 y 41.80, $W^{(1)}$ entre .370 y 7.565, y $W^{(3)}$ entre 8.33 y 9.248; ninguno fue identificado como evidencia sustancial de dependencia local entre los ítems.

Monotonicidad. En la Figura 1 se muestra el patrón incremental de las opciones de respuesta, y del puntaje de los ítems. Las curvas menos diferenciadas con las correspondientes a las primeras opciones de respuesta, y todos los ítems mantienen un incremento constante y prácticamente lineal.

Tabla 2.

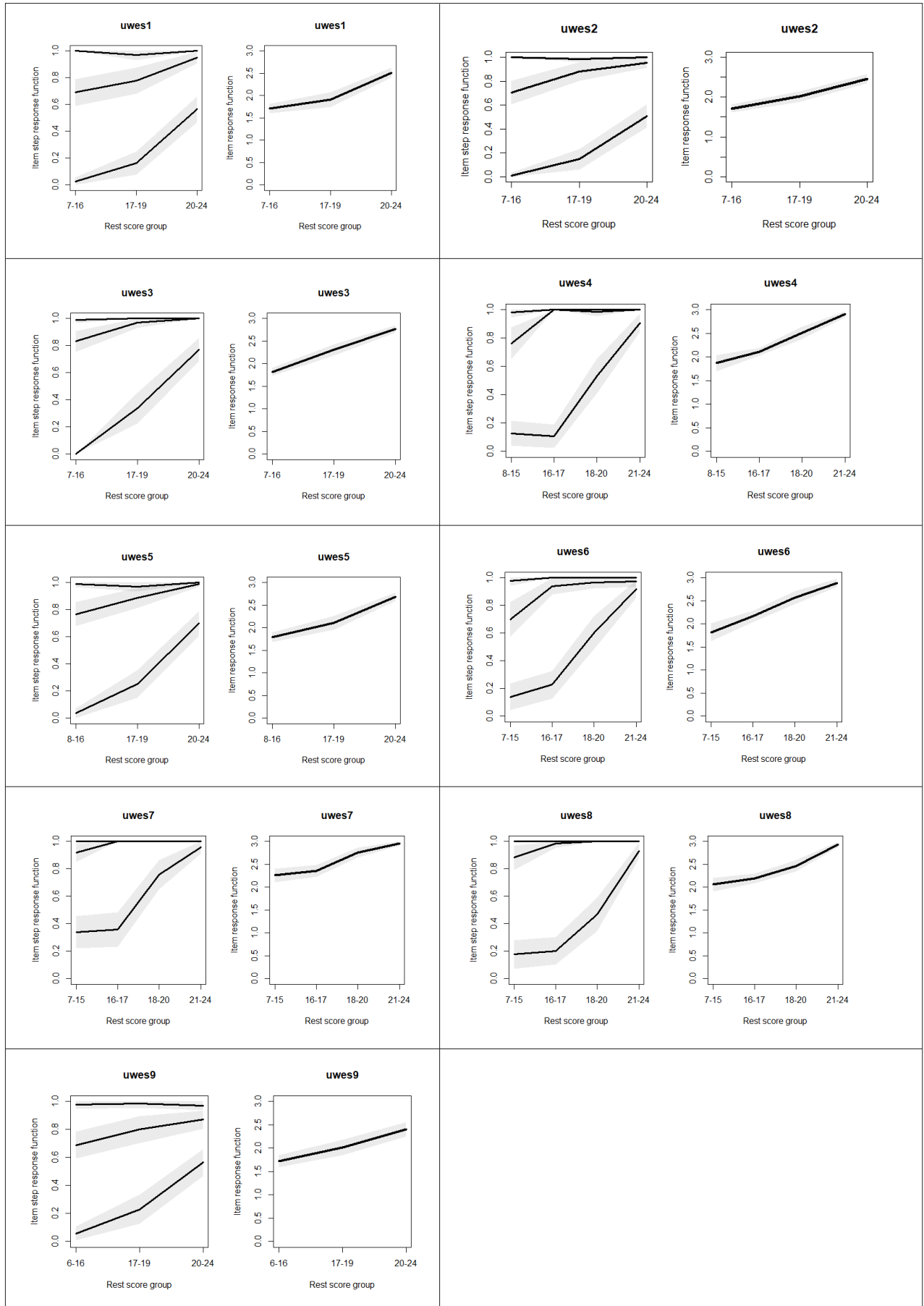
Escalamiento Mokken: Procedimiento automatizado de selección de ítems (AISP)

	AISP (método normal)					Escalamiento			Homogeneidad		
	.30	.40	.50	.60	.70	(Hi)			monotónica (MHM)		
						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	-	-	-

Nota. AISP: identificación de escalas derivadas de los ítems; 0 significa que el ítem no se agrupa en ninguna. Hi: escalabilidad para cada ítem. ^a Ítems de la versión abreviada UWES-3

Figura 1.

Monotonicidad de los ítems: opciones de respuesta y puntaje del ítem.



Modelamiento lineal (SEM)

En la versión UWES-9 (Tabla 3), el modelo de tres factores correlacionados se ajustó satisfactoriamente: WLSMV- $\chi^2 = 122.636$, $gl = 2$ ($p < .01$); CFI = .989, TLI = .984, SRMR = .076; sin embargo, las correlaciones interfactoriales fueron elevadas, sugiriendo falta de discriminación conceptual entre ellas. El modelo unidimensional también fue satisfactorio, WLSMV- $\chi^2 = 316.665$, $gl = 27$ ($p < .01$); CFI = .968, TLI = .958, SRMR = .122, aunque ligeramente bajo. El ajuste del UWES-3 fue completamente satisfactorio: WLSMV- $\chi^2 = 638.132$, $gl = 3$ ($p < .01$); CFI = 1.000, TLI = 1.000, SRMR = .000. Las cargas factoriales en todos los modelos fueron elevados ($> .50$) y estadísticamente significa-

tivos. El ítem 3 en el modelo del UWES-3 mostró carga superior a 1.0, pero no fue necesariamente un caso *Heywood* debido a su magnitud. No se detectaron modificaciones al modelo unidimensional del UWES-9, y no fueron explorados.

Se observa también que los ítems correspondientes al UWES-3 constantemente fueron los de mayor carga factorial comparados con el resto de los ítems, en los modelos de tres dimensiones y en el de una dimensión (Tabla 3). Los umbrales (la separación de las opciones en el continuo latente del puntaje del UWES-3) fueron irregulares en su separación, y el ítem 8 mostró que la última opción de respuesta no fue funcional debido a la ausencia de respuestas, y el umbral no fue estimado.

Tabla 3.
Modelamiento SEM del UWES-9 y UWES-3

	UWES – 9						UWES – 3			
	3 dimensiones			Unidimensional			FT	Umbrales		
	VI	DED	AB	R ²	FT	R ²		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				
Correlaciones										
VI	1.0									
DED	.717	1.0								
AB	.470	.707	1.0							
Confiabilidad										
α					.877		.703			
(IC 95%)					(.84, .89)		(.62, .76)			
ω					.871		.754			
(IC 95%)					(.83, .89)		(.70, .81)			

Nota. VI: Vigor. DED: Dedicación. AB: Absorción. FT: factor total. Ítems de la versión abreviada del UWES. α y ω : coeficientes de confiabilidad. τ ítems de la versión abreviada UWES-3

Confiabilidad

Desde el modelo MSA, la confiabilidad MS para el UWES-3 fue .892, un nivel que puede considerarse alto. Asimismo, las confiabilidades alfa (α) y omega (ω) fueron moderadamente altas ($\geq .70$), y se observa leve discrepancia entre ambas (Tabla 3, parte inferior).

Equivalencia entre versiones

La correlación no ajustada ($r = .931$, $p < .01$) y ajustada ($r = .824$, $p < .01$) entre UWES-9 y UWES-3 fue elevada, indicando muy similar ordenamiento de sujetos desde ambos puntajes. El acuerdo clasificatorio fue moderadamente alto, $AC_1 = .611$ (IC 95% = .537, .687), indicando la tendencia a diferenciar sujetos cuando son clasificados en cuartiles. Respecto a equivalencia de la validez o discriminación de constructo con el síndrome de quemarse por el trabajo, las correlaciones del UWES-9 (puntaje único basado en el modelo unidimensional;

ver párrafos arriba) y UWES-3 obtenidas mostraron coherencia teórica (Tabla 4). La correlación con la dimensión de Ilusión por el Trabajo del CESQT-PE fue positiva y alta, mientras que con el resto de los puntajes fue de dirección negativa y bajas o moderadamente bajas, como se esperaba. Respecto la diferencia correlacional, ambas versiones produjeron magnitudes correlacionales similares; la magnitud de estas diferencias estimadas fue pequeña, porque: a) entre 8% y 11% más bajo respecto a la versión UWES-9, b) los intervalos de confianza para las diferencias correlacionales (Δ_r) incluyeron el cero (excepto en Ilusión por el trabajo), y c) la diferencia estandarizada (q) fue entre .020 y .083, es decir, de magnitud insustancial. Considerando la edad y el sexo de los participantes, los puntajes de ambas versiones correlacionaron cero en la población, y la diferencia correlacional fue insustancial (q alrededor de cero) (Tabla 4).

Tabla 4.
Correlaciones entre versiones del UWES con CESQT

	Correlaciones		Diferencia correlacional		M	DE
	UWES-9	UWES-3	Δ_r (IC 95%)	q		
Sexo	-.082	-.109	(-.018, .072)	.027	-	-
Edad	.068	.035	(-.012, .078)	.033	-	-
CESQT						
Ilusión por el trabajo	.618**	.564**	(.017, .095)*	.083	3.33	.653
Indolencia	-.245**	-.226**	(-.064, .025)	.020	.67	.555
Desgaste	-.384**	-.413**	(-.013, .072)	.034	1.75	.976
Culpa	-.188**	-.168**	(-.065, .025)	.020	.62	.620
M	29.66	9.85	-	-	-	-
DE	4.07	1.47	-	-	-	-

Nota. Δ_r : diferencia de correlaciones dependientes. q : magnitud de la diferencia. CESQT: dimensiones del Cuestionario para la Evaluación del Síndrome de Quemarse por el Trabajo. * $p < .05$; ** $p < .01$

Discusión

El objetivo del presente estudio fue analizar las propiedades psicométricas de la versión ultra-breve UWES-3 y analizar su equivalencia con la UWES-9 en una muestra de docentes universitarios mexicanos. La estrategia metodológica comprendió una exploración de la dimensionalidad y un examen de las propiedades estructurales, mediante estimaciones paramétricas y no paramétricas, así como un análisis de equivalencia de los patrones de correlaciones directos y ajustados, tanto entre ambas versiones (UWES-3 y UWES-9) como con las dimensiones del CESQT, el sexo y la edad.

Dicha estrategia metodológica elegida tiene ventajas sobre estudios previos, específicamente en la inclusión de estimaciones de correlación ajustadas, omitidas en el estudio de [Schaufeli et al. \(2017\)](#) y la estimación de resultados no paramétricos junto con los resultados paramétricos, omitidas en el estudio de [Domínguez-Lara et al. \(2021\)](#). Respecto a las correlaciones ajustadas, el control de la varianza de error mediante la corrección de [Levy \(1968\)](#) aplicada a la correlación entre ambas versiones, es una práctica que se recomienda para obtener conclusiones válidas de equivalencia ([Merino-Soto, & Angulo-Ramos, 2013](#); [Smith, McCarthy, & Anderson, 2000](#)), debido a que se tienden a obtener tamaños correlacionales espurios sin esta corrección. Por otro lado, la aplicación del modelamiento no paramétrico generalmente es considerado como un precursor de modelos paramétricos, debido que no requiere una conocida distribución del constructo (eg, distribución normal) ([Mokken, 1971](#); [Van der Ark, 2012](#)).

Los análisis dimensionales y estructurales paramétricos y no paramétricos coincidieron en la confirmación de una estructura unidimensional del UWES-9, donde justo los 3 ítems de la versión breve presentaron las cargas factoriales más altas y un ajuste paramétrico totalmente satisfactorio. Esto confirma la unidimensionalidad esencial del UWES-3, con alta validez de los ítems en relación con su constructo. Por otro lado, la consistencia interna fue aceptable para ambas versiones, aunque superior para el UWES-9. El decremento de la confiabilidad de los puntajes del UWES-3 es una consecuencia del número de ítems reducido y su

consecuente restricción de la varianza del constructo, pero no es necesariamente un problema en una visión relativa, porque la magnitud apropiada o razonable de la confiabilidad de un puntaje está condicionado por el uso de este puntaje y el contexto de aplicación (American Educational Research Association (AERA), American Psychological Association (APA) & National Council for Measurement in Education (NCME), 2014). Por lo tanto, con una magnitud de confiabilidad alrededor de .75 de los puntajes del UWES-3, su uso parece adecuarse a evaluaciones de tamizaje, monitoreo, vigilancia psicosocial y descripciones de grupos. Para usos adicionales que requieran mayor precisión del puntaje, la versión de UWES-9 puede estar mejor indicada. Como nota adicional, la diferencia entre el coeficiente alfa y omega del UWES-3 es consecuencia de la distancia de la carga factorial del ítem 3 respecto a los otros dos ítems. Una implicación es que, si esto se mantiene en otros estudios, entonces el coeficiente adecuado para estimar la confiabilidad es ω , no α .

Uno de los hallazgos del estudio fue la magnitud de los coeficientes de escalabilidad H en el grupo de mujeres, comparado con los varones. Los coeficientes de escalamiento H son interpretables como coeficientes de discriminación de un ítem respecto al puntaje de la media ([Van Abswoude, Van der Ark, & Sijtsma, 2004](#)), y parcialmente se refiere también a la calidad de medición de ítem. Esto implica que las propiedades de escalamiento y de calidad general del UWES-3 en mujeres, no son tan fuertes como en el caso de los varones. Resultados como los obtenidos aquí, en que las diferencias de grupo son aparentes en la calidad de medición, requieren más investigación ([Wind, 2017](#)). En el funcionamiento de diferencial de ítems, dentro del marco *Mokken Scale Analysis* (MSA), la intersección de curvas características de las opciones de respuesta puede estar detrás de estas diferencias, demostrando una menor capacidad de discriminación de diferencias individuales en mujeres. Es probable que estas diferencias sean idiosincrásicas a la muestra de participantes, y por lo tanto no replicables, sin embargo, esto alerta sobre la necesidad de incluir otros indicadores en una condición de mayor tamaño muestral.

En cuanto a la equivalencia de las versiones UWES-3 y UWES-9, las correlaciones entre ambas versiones presentaron coeficientes superiores a $r = .82$, con un acuerdo clasificatorio moderadamente alto y los patrones de correlaciones con el CESQT y las variables demográficas (sexo, edad) fueron prácticamente similares tanto teórica como empíricamente, en la misma dirección y con diferencias de tamaño estadísticamente insustanciales. Ello que lleva a la conclusión que el UWES-3 es psicométricamente equivalente al UWES-9 para la evaluación del engagement laboral en profesores universitarios.

Resulta relevante señalar que la versión ultra-breve de 3 ítems (UWES-3) representa en cada ítem los tres mismos aspectos conceptuales importantes del engagement que fueron acuñados desde su origen: vigor, dedicación y absorción, por lo que no se pone en riesgo la validez de contenido, ni se pierde representación teórica construida en sus raíces. En general, no hay pérdidas significativas de información o varianza con respecto a las versiones del UWES con más ítems, lo que implica beneficios prácticos implícitos.

Respecto a ello, actualmente existe una tendencia de investigar la viabilidad de medidas ultra-breves debido a los beneficios prácticos que esto tiene, las ventajas son múltiples e incluyen entre otras: tomar menos tiempo de participantes e instituciones en el llenado de cuestionarios, disminuir las tasas de deserción por cuestionarios largos, evitar la fatiga o monotonía de los respondientes (lo que al mismo tiempo asegura la validez de las mediciones), y no menos importante, se facilita su inclusión en estudios epidemiológicos grandes o encuestas nacionales midiendo los mismos constructos sin muchos ítems. Estos beneficios pueden hacerse extensivos a la versión breve UWES-3 analizada en este estudio y la opción de una evaluación más práctica y sencilla del engagement laboral en profesores universitarios, aunque parecen necesarios más estudios que confirmen la tendencia encontrada hasta ahora.

Las limitaciones del estudio pudieran identificarse en el tamaño y la estrategia no aleatoria de la muestra, la elección de pocas variables para la discriminación (dimensiones CESQT, edad, sexo) y la implícita falta de independencia del uso del

UWES-9 y UWES-3 en el mismo estudio, lo que dificulta entender su verdadera confiabilidad y validez de forma verdaderamente independiente (Schaufeli et al., 2017). Sin embargo, las estimaciones no paramétricas y robustas utilizadas pueden atenuar en alguna medida los problemas de muestreo. Asimismo, la elección del burnout o síndrome de quemarse por el trabajo, medido mediante el CESQT, como criterio de análisis de equivalencia pudo ser suficiente dada la divergencia teórica esperada entre ambos constructos, considerando que el engagement nació y cobró identidad como justo un opuesto del Burnout. Corresponde a futuros estudios analizar el patrón de correlaciones con otras variables, en otras muestras, utilizando de forma conjunta e independiente el UWES-9 y el UWES-3 para confirmar tal equivalencia en distintos escenarios. Entretanto, sirva este estudio como una aportación inicial en este sentido, en beneficio de la medición del *engagement* en la fuerza laboral docente latinoamericana.

Referencias

- 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., Domínguez-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. *PLoS ONE*, 10(4), e0121945. <http://doi.org/10.1371/journal.pone.0121945>
- Domínguez-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](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. *Psicología 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 Association (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>