Zuleima Santalla-Banderali1 and Jesús M. Alvarado2

**Resumen**

Se ha evaluado la validez factorial of the Spanish short version of Utrecht Work Engagement Scale (UWES-9) en su adaptación al Ecuador, así como su validez predictiva respecto al desempeño laboral autoevaluado. Participaron 229 trabajadores de instituciones educativas de Ecuador. En un análisis de comparación de modelos, se observó una excelente bondad de ajuste del modelo unidimensional que no era mejorada por modelos más complejos, lo que justifica su uso como instrumento unidimensional del work engagement (WE). Ahora bien, al analizar los patrones de correlación de la puntuación global y de las dimensiones del WE respecto a the task and contextual performance, and counterproductive behaviours, se concluye que, a pesar de observarse un buen ajuste del modelo unidimensional, la propuesta teórica de tres factores es sustantivamente superior en cuanto a que mantiene una validez predictiva diferencial para cada dimensión teórica.

**Introduction**

En la búsqueda de la comprensión del comportamiento de las personas en el ámbito laboral, el foco de los académicos ha ido pasando a estar dirigido al comportamiento organizacional positivo (Balducci et al., 2010) y los estados psicológicos positivos experimentados por los trabajadores (Schaufeli, 2013). En este contexto, uno de los aspectos que ha recibido especial atención es el work engagement. Así lo evidencian los resultados obtenidos por Motyka (2018) en su revisión sistemática de la literatura, en la que analizó 71 investigaciones sobre engagement publicadas en 25 países entre el 2002 y el 2017, observando que el 70% de dichas investigaciones habían sido publicadas entre el 2014 y el 2018.

Ahora bien, el auge del interés por el estudio del work engagement no se ha dado por igual en todos los países y continentes. En este sentido, Motyka (2018) encontró que el 20% de los estudios incluidos en su revisión se habían realizado en Estados Unidos de Norteamérica y en Holanda, y que el 46% provenía de Europa; no habiendo ningún estudio realizado en países de Latinoamérica.

La relevancia del work engagement se evidencia en varios modelos teóricos que incluyen a este constructo como un mediador de la relación entre otras variables y el desempeño laboral. Por ejemplo, Bakker (2011) propuso que los recursos laborales (e.g. el apoyo social de colegas y supervisores, el feedback respecto al desempeño, la variedad de habilidades requeridas por las tareas a realizar y la autonomía en el trabajo), y los recursos personales (e.g. el optimismo, la autoeficacia, la resiliencia y la auto-estima) dan inicio a un proceso motivacional que redunda en un incremento del work engagement, lo cual a su vez da como resultado un mejor desempeño. De modo similar, Christian et al. (2011) plantearon en su modelo conceptual que el engagement media las relaciones entre las características del trabajo, el liderazgo, las características disposicionales, y el task and contextual performance.

Adicionalmente, hay amplia evidencia empírica que muestra que entre el work engagement y el job performance existe una relación positiva y estadísticamente significativa (Afacan-Findikli, 2015; Breevaart et al., 2015; Christian et al., 2011; Dalal et al., 2012; Gutermann et al., 2017; Kim et al., 2013; Lorente et al., 2014; Lupano et al., 2018; Motyka, 2018; Pongton & Suntrayuth, 2019; Qodariah et al., 2019; Rich et al., 2010). De esta forma, el work engagement es considerado un antecedente clave del comportamiento de los trabajadores (Bakker et al., 2011; Bothma & Roodt, 2012), que contribuye a la ventaja competitiva de las organizaciones, aumentando su rendimiento financiero y reputación (Karanges et al., 2014; Shuck & Reio, 2011; Tampubolon, 2016).

A pesar de lo anterior, el work engagement es un constructo complejo y aún hoy en día no hay consenso en cuanto a su conceptualización (Chaudhary et al., 2012; Wefald et al., 2012), si bien varias de las definiciones que se han propuesto podrían considerarse complementarias.

**Conceptualización del work engagement**

La complejidad del work engagement (WE) queda de manifiesto desde el momento en que se ha discutido incluso si es un constructo en sí mismo o se solapa con otros constructos ya utilizados en el ámbito organizacional. En este sentido, la evidencia empírica muestra de forma sistemática que the WE es un constructo distinto a otros como organisational commitment, job involvement, job satisfaction, organizational citizenship behaviour and extra-role behavior (Bakker et al., 2011; Kataria et al., 2013; Saks, 2006). De igual forma, actualmente también parece superada la diatriba sobre si el WE es o no el polo opuesto del burnout. Schaufeli (2013) y Schaufeli and Salanova (2011) explican que estos dos constructos son independientes, poseen estructuras distintas y explican aspectos diferentes de la relación que establecen los empleados con sus trabajos (Schaufeli et al., 2002). Así, aunque entre ambos existe una relación inversa (Schaufeli, 2013), esta dista mucho de ser perfecta, de forma que el hecho de que una persona no experimente burnout no significa que se sienta engaged and viceversa (Schaufeli & Salanova, 2011).

La discusión que aún sí se mantiene es la relativa a la uni o multidimensionalidad del constructo. En relación con ese punto, autores como Britt (1999, as cited in Welfald et al., 2012) and the Gallup Organization coinciden en concebir the WE como un constructo unidimensional; pero, lo definen de forma distinta. Para Britt, the WE es un estado motivacional que hace referencia a en qué medida al empleado le importa su desempeño laboral, sintiéndose responsable y comprometido con dicho desempeño (Welfald et al., 2012); mientras que, para the Gallup Organization the WE alude a la participación, satisfacción y entusiasmo del individuo por su trabajo (Schaufeli, 2013; Shuck, 2011).

A diferencia de lo anterior, Kahn (1990), reconocido como el primero en aplicar la noción de engagement al contexto laboral, lo concibió como un constructo multidimensional conformado por tres dimensiones: (a) cognitive engagement, (b) emotional engagement, and (c) behavioral engagement. Este autor definió el WE como un estado psicológico en el que los empleados "incorporan" su yo personal a sus actuaciones laborales, invirtiendo energía personal y experimentando una conexión emocional con su trabajo; de forma que están conectados con los roles de su trabajo física, cognitiva y emocionalmente (Bakker, 2011; Schuck, 2011, 2013; Shuck & Reio, 2011), dedican sus recursos personales a la realización de las tareas asociadas al rol que desempeñan dentro de la organización (Christian et al., 2011), y focalizan su energía en los objetivos organizacionales (Bakker, 2011).

Para Kant, the cognitive engagement corresponde a la evaluación que hacen los empleados sobre si el trabajo que realizan es significativo y seguro, y si cuentan con los recursos adecuados para realizarlo (Shuck & Reio, 2011). The emotional engagement alude al vínculo emocional que la persona siente con su lugar de trabajo y representa la voluntad del individuo de involucrar sus recursos personales en su trabajo (Shuck & Reio, 2011). Finalmente, the behavioral engagement es la intención de actuar en aras de un mejor rendimiento, siendo la manifestación explícita del cognitive and emotional engagement (Shuck & Reio, 2011).

La visión del WE como un constructo multidimensional es compartida por Schaufeli y sus colegas, quienes lo definen como un estado psicológico positivo, satisfactorio y activo asociado con la realización de las actividades laborales diarias, caracterizado por el vigor, la dedicación y la absorción (Bakker, 2011; Salanova-Soria & Schaufeli, 2004; Schaufeli, 2013, 2017; Schaufeli et al., 2002), siendo the vigour and dedication the core dimensions (Schaufeli & Salanova, 2011). De acuerdo con Schaufeli (2013, 2017), esta noción es congruente con la de Kant (1990), en el sentido de que hay una equivalencia entre physical-energetic and vigor, emotional energy and dedication, and a cognitive component and absorption.

El vigor hace referencia a en qué medida el trabajador experimenta altos niveles de energía y resistencia física y mental mientras trabaja (Bakker, 2011; Bakker et al., 2011, 2012; Schaufeli, 2013); su deseo de invertir esfuerzo en la labor que realiza y de persistir aún en momentos de dificultad (Salanova-Soria & Schaufeli, 2004; Schaufeli, 2013, 2017; Schaufeli et al., 2002).

La dedication hace referencia al grado con el cual la persona se siente involucrada e identificada con el trabajo que realiza y experimenta un sentido de importancia, entusiasmo, inspiración, orgullo y desafío (Bakker, 2011; Bakker et al., 2011, 2012; Salanova-Soria & Schaufeli, 2004; Schaufeli, 2013, 2017; Schaufeli et al., 2002).

Finalmente, la absorption alude a en qué medida la persona está tan concentrada y absorta en su trabajo (Bakker, 2011; Bakker et al., 2011, 2012; Schaufeli, 2013) que el tiempo le pasa rápidamente (Bakker, 2011; Schaufeli, 2013), y tiene dificultades para separarse del trabajo (Salanova-Soria & Schaufeli, 2004; Schaufeli, 2013, 2017; Schaufeli et al., 2002; Simpson, 2009).

Para Schaufeli y sus colegas, the WE es relativamente estable en el tiempo y no se focaliza en un evento, objeto, persona o conducta particular (Salanova-Soria & Schaufeli, 2004; Schaufeli & Salanova, 2011; Schaufeli et al., 2002; Simpson, 2009). No obstante, esto no significa que no pueda cambiar dependiendo de las características del trabajo (Schaufeli & Salanova, 2011) y lo que suceda a lo largo del día (Bakker, 2011). De hecho, algunos investigadores citados en Breevaart et al. (2012) han medido diariamente el constructo, confirmando que the WE reportado por una persona en particular puede fluctuar a lo largo del tiempo.

La concepción del WE propuesta por Schaufeli y sus colegas es la adoptada mayoritariamente por los investigadores, según los resultados de la revisión de la literatura realizada por Motyka (2018), y es la noción que se asumió en el presente estudio.

**Medida del work engagement**

Si bien se han desarrollado diversos instrumentos dirigidos a medir the WE, la medida más ampliamente usada (Rich et al., 2010; Schaufeli, 2013; Shuck, 2011; Welfald et al., 2012) es the Utrecht Work Engagement Scale (UWES) creada por Schaufeli y sus colegas sobre la base de su conceptualización del engagement como un constructo multidimensional; instrumento el cual es incluso considerado la medida estándar de este constructo (Kataria et al., 2013; Kulikowski, 2017; Lovakov et al., 2017; Rodríguez-Montalbán et al., 2014; Souza-Vázquez et al., 2017; Willmer et al., 2019). De hecho, en la revisión sistemática de Motyka (2018) se halló que en el 82% de las publicaciones analizadas, the WE había sido medido usado alguna de las versiones de las UWES.

Originalmente the UWES estaba compuesta por 24 ítems (Schaufeli et al., 2002, 2006); pero, luego de varias evaluaciones psicométricas y con objeto de lograr una medición más parsimoniosa, los autores redujeron su longitud a 17 ítems (Kataria et al., 2013; Rodríguez-Montalbán et al., 2014; Schaufeli et al., 2006): seis que medían la dimensión vigour, cinco para la dimensión dedication and seis para la dimensión absorption. No obstante, pueden encontrarse trabajos con una versión de la escala con 15 ítems, cinco por cada una de las dimensiones, debido a los resultados que mostraban que dos de los ítems de la versión UWES-17 eran inconsistentes (Salanova et al., 2000).

Posteriormente, Schaufeli et al. (2006) desarrollaron una versión de nueve ítems (UWES-9), tres por cada una de las dimensiones, que fue validada en 10 países con una muestra de más de 14.000 personas. Puede afirmarse que esta es la versión usada con mayor frecuencia, pues según Motyka (2018), esta versión fue la empleada en el 68% de los artículos analizados en su revisión de la literatura.

La UWES-9 tiene propiedades psicométricas equivalentes a sus versiones más largas (Chaundhary et al., 2012; Kataria et al., 2013; Schaufeli et al., 2006), observándose que incluso tiene una utilidad práctica y una validez de constructo superiores a la UWES-17 (Schaufeli et al., 2006; Seppälä et al., 2009). De hecho, en la revisión realizada por Kulikowski (2017), de 21 estudios en los que se evaluó la validez de constructo de la UWES usando análisis factorial confirmatorio (CFA), se halló que en ocho de los nueve estudios en los que se comparó la UWES-17 con la UWES-9, esta última versión mostró un mejor comportamiento psicométrico. Además, las puntuaciones en la UWES-9 son más estables en el tiempo y explican en torno al 80% de la variación en las puntuaciones en la UWES-17 (Kulikowski, 2017).

**Propiedades psicométricas de la UWES-9**

The UWES-9 presenta altos índices de consistencia interna, reportándose coeficientes Alpha de Crombach para la escala completa superiores a 0.80 (Balducci et al., 2010; Chaundhary et al., 2012; Kataria et al., 2013; Klassen et al., 2012; Schaufeli et al., 2006; Villotti et al., 2014; Welfald et al., 2012; Willmer et al., 2019; Yusoff et al., 2013); y valores α entre 0.59 y 0.92 para las distintas dimensiones (Balducci et al., 2010; Chaundhary et al., 2012; Hallberg & Schaufeli, 2006; Kataria et al., 2013; Klassen et al., 2012; Lovakov et al., 2017; Rodríguez-Montalbán et al., 2014; Schaufeli et al., 2006; Villotti et al., 2014; Welfald et al., 2012).

Por otra parte, en cuanto a la estructura multidimensional de la UWES, hay amplias discrepancias en los resultados obtenidos, lo cual pone en duda la generalizabilidad de la noción del work engagement como un constructo conformado por las tres dimensiones planteadas por Schaufeli y sus colegas. En este sentido, Schaufeli et al. (2006), Villotti et al. (2014) and Willmer et al. (2019) obtuvieron correlaciones entre las tres dimensiones mayores a 0.70. Breevaart et al. (2012), De Bruin and Henn (2013), Hallberg and Schaufeli (2006), and Souza-Vázquez et al. (2017) encontraron correlaciones superiores a 0.80; y Balducci et al. (2010), Chaundhary et al. (2012), Gómez-Garbero et al. (2019), Lovakov et al. (2017), and Schaufeli et al. (2006) reportaron correlaciones superiores a 0.90. Rodríguez-Montalbán et al. (2014) hallaron que, en promedio, las tres sub-escalas de la UWES comparten el 57% de la varianza, lo cual nuevamente indica que las dimensiones teóricas de la UWES están altamente correlacionadas y sugieren unidimensionalidad.

Adicionalmente, en Chile y usando análisis factorial exploratorio (EFA), Müller et al. (2013) obtuvieron dos factores. A diferencia de esto, también empleando EFA, Chaundhary et al. (2012) en la India, Souza-Vázquez et al. (2017) en Brazil, y Willmer et al. (2019) en Suecia encontraron una solución unidimensional; pero, el confirmatory factor analysis (CFA) mostró que el modelo de tres factores tenía índices de ajuste ligeramente superiores a la solución unifactorial en los casos de la India y Brazil.

El modelo de tres factores también ha mostrado índices de ajuste aceptables en Uruguay (Gómez-Garbero et al., 2019), superiores al unifactorial en Italia (Balducci et al., 2010), en Holanda (Balducci et al., 2010; Breevaart et al., 2012) y con la muestra multinacional usada por Schaufeli et al. (2006), y mejores al de uno y dos factores (dedicación+vigor, y absorción) en Pakistan (Yusoff et al., 2013). Ahora bien, en estudios como el de Balducci et al. (2010) los resultados indicativos de un buen ajuste del modelo de tres factores se obtuvieron solo cuando se modificó para permitir the error covariance between cuatro pares de ítems en el caso de la muestra de Italia, y entre dos pares de ítems en la muestra de Holanda. Esta correlación entre errores sugiere un posible solapamiento de contenido entre los ítems (o problema de redundancia).

En contraposición con lo anterior, en investigaciones como la de Souza-Vázquez et al. (2017) en Brasil, Hallberg and Schaufeli (2006) en Suecia, y Villotti et al. (2014) en Italia los resultados han mostrado diferencias no estadísticamente significativas entre los modelos unidimensional y de tres factores en cuanto a sus índices de ajuste. Adicionalmente, Rodríguez-Montalbán et al. (2014) en Puerto Rico y Lovakov et al. (2017) en Rusia encontraron que el modelo de tres factores presentaba un ajuste ligeramente superior al unidimensional, aunque este último también tenía índices de ajuste aceptables cuando los modelos se modificaban para permitir the error covariance between three pairs of items (Lovakov et al., 2017), con lo que se puede controlar la posible redundancia de contenido.

Los cuestionamientos sobre la estructura multidimensional de la UWES aumentan al considerar que hay estudios en los que, a diferencia de los anteriores, lo que se ha observado es una superioridad del modelo unidimensional. Por ejemplo, trabajando con muestras de Australia, Canadá, China, Indonesia and Oman, Klassen et al. (2012) encontraron que, considerando conjuntamente los datos de las muestras de los cinco países y cuando los modelos uni y trifactorial se modificaron para permitir the error covariance between ítems, los índices de ajuste del modelo de un factor eran ligeramente mejores que los del modelo de tres factores. Así mismo, Fong and Ho (2015), usando Bayesian structural equation modeling (BSEM) and especificando informative residual covariance priors, hallaron que si bien los tres modelos probados (one-factor, three-factor and partial bi-factorial model) presentaban índices de ajuste adecuados, el unidimensional BSEM tenía un BIC significativamente más bajo, indicando un mejor ajuste que los modelos de tres factores y partial bi-factor BSEM. Sin embargo, De Bruin and Henn (2013) hallaron en Sudáfrica que el modelo con los mejores índices de ajustes era un partial bifactorial model which specifies a general factor que los autores denominaron “work engagement” y que explicó el 89% de la varianza, y dos factores específicos: vigor y absorción.

Finalmente, también hay resultados que muestran el inadecuado ajuste de los diversos modelos probados: un factor, dos factores y tres factores (Kataria et al., 2013; Wefald et al., 2012; Willmer et al., 2019). En el caso del estudio de Klassen et al. (2012) esto fue así cuando se evaluaron los modelos sin permitir la correlación entre los errores. En el caso del estudio de Fong and Ho (2015) ninguno de los modelos probados (un factor, tres factores y partial bifactorial) presentó índices de ajuste adecuados cuando el ajuste se evaluó usándose the traditional ML approach of CFA. Finalmente, Hallberg and Schaufeli (2006) and Villotti et al. (2014) encontraron que, tanto para el modelo de tres factores como para el unidimensional, los índices de ajuste RMSEA y χ2 indicaban un ajuste inaceptable.

La discrepancia de resultados en cuanto a la estructura de la UWES también ha sido puesta de manifiesto por Kataria et al. (2013) quienes hallaron que el modelo unifactorial tenía un mejor ajuste en el 33.33% de los 10 estudios analizados; pero, en otro 33.33% se halló un mejor ajuste de la solución trifactorial, y en el 20% se observó que ninguna de las dos soluciones presentaba índices de ajuste adecuados. Más recientemente, Kulikowski (2017) revisó 21 estudios en los que se evaluó la estructura factorial de la versión original y la versión abreviada de la UWES, empleándose CFA, y confirmó la inestabilidad en la estructura factorial de la escala: en el 23.81% de los estudios se encontró la estructura trifactorial original; pero, en otro 23.81% se halló una estructura unidimensional, y en el 38.09% los resultados mostraron que ambos modelos (unifactorial y tres factores) eran equivalentes.

Esta variabilidad de resultados evidencia que la UWES-9 puede tener un problema en cuanto a la invarianza de su estructura factorial que requiere seguir siendo investigado (Kulikowski, 2017) a fin de alcanzar una mejor comprensión de la misma. Ampliar el conocimiento en cuanto a la estructura factorial de la UWES-9 es relevante no solo desde el punto de vista teórico sino práctico. Concretamente, confirmar que la estructura de la escala para la que se obtienen mejores índices de ajuste difiere al trabajarse con muestras distintas, implicaría que antes de emplear esta escala sería necesario chequear su comportamiento en la población concreta con la que se pretenda trabajar; de forma que, en función de lo hallado en cada caso particular, tomar decisiones apropiadas en cuanto a si se deben usar e interpretar las puntuaciones de cada una de las sub-escalas por separado, o lo más apropiado es usar e interpretar una puntuación total única. Por otra parte, la confirmación de un problema de invarianza conlleva a que resulte inapropiado comparar resultados obtenidos en estudios en los que se ha trabajado con muestras diferentes para las que se han hallado configuraciones distintas.

Continuar con la investigación sobre el comportamiento psicométrico de la UWES-9 es especialmente relevante en países latinoamericanos en los que es escasa la evidencia científica disponible sobre este aspecto (Rodríguez-Montalbán et al., 2014) y, por ende, sobre su aplicabilidad concretamente en Ecuador. Como señalan Klassen et al. (2012) “measures developed in particular settings must be carefully examined across cultural settings before they can be reliably used beyond those settings” (p. 318).

Partiendo de lo anterior, el presente estudio tuvo como objetivo, por una parte, conocer cuál es la estructura factorial del UWES-9 más adecuada para la medida del constructo en su adaptación al Ecuador; y, por otra parte, conocer su validez predictiva respecto a la medida del desempeño laboral autoevaluado, variable esta que, como se indicó previamente, es la que con mayor frecuencia se pretende predecir a partir del work engagement.

Concretamente, usando Confirmatory Factor Analysis, se pusieron a prueba los siguientes modelos de medida:

* Tres factores correlacionados, basado en los estudios que muestran altas correlaciones entre las tres dimensiones propuestas por Schaufeli y sus colegas (Balducci et al., 2010; Breevaart et al., 2012; Chaundhary et al., 2012; De Bruin & Henn, 2013; Gómez-Garbero et al., 2019; Hallberg & Schaufeli, 2006; Lovakov et al., 2017; Rodríguez-Montalbán et al., 2014; Schaufeli et al., 2006; Souza-Vázquez et al., 2017; Villotti et al., 2014; Willmer et al., 2019).
* Dos factores correlacionados (dedicación + vigor, y absorción), considerado previamente por Wilmer et al. (2019) y Yusoff et al. (2013).
* Un modelo unidimensional basado en los resultados de autores como Fong and Ho (2015), Klassen et al. (2012) y Lovakov et al. (2017), o esencialmente unidimensional: (a) a classical bifactor model (un factor general de work engagement y tres factores específicos: absorción, dedicación y vigor), que permite evaluar la validez de las puntuaciones de factores específicos, controlando la varianza debida al factor general (Ondé et al., 2021); y, (b) a S-1 bifactor model or partial bifactor (De Bruin & Henn, 2013; Fong & Ho, 2015), en el que se utiliza uno de los factores específicos (i.e. factor de referencia) como factor general (en nuestro caso, el factor “dedicación”) y se analiza el resto de los factores específicos controlados por la varianza común del factor de referencia (Ondé et al., 2021).

**Method**

**Sample**

En el estudio participaron 229 trabajadores de instituciones educativas privadas de Guayaquil-Ecuador (73.4% mujeres) con una edad promedio de 36.04 años (SD = 9.38). En cuanto a la categoría profesional, el 67.6% de la muestra eran profesionales de alta cualificación y el 13.1% era personal administrativo o auxiliar. El 68.1% tenía título universitario y el 16.2% tenía alguna titulación de postgrado. La mayoría de los participantes (72.1%) tenía más de cinco años de experiencia laboral, el 96.8% tenía un contrato fijo indefinido, y el 95.4% trabajaba jornada completa. Finalmente, el 43.7% de los sujetos llevaba entre 1 y 5 años trabajando en su organización actual y el 28.4% llevaba trabajando en la organización más de cinco años.

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

**Instruments**

***Work engagement***

Para medir el work engagement se empleó the Utrech Work Engagement Scale (UWES-9) validada por Schaufeli et al. (2006), con los ítems en castellano de la versión UWES-17 de Salanova et al. (2000). En esta escala se pide a las personas que indiquen con qué frecuencia se han sentido del modo expresado en cada uno de los ítems, usando una escala Likert de seis intervalos (1 = nunca/casi nunca a 6 = casi siempre/siempre). Al igual que lo reportado por Balducci et al. (2010), Chaundhary et al. (2012), Hallberg & Schaufeli (2006), Kataria et al. (2013), Klassen et al. (2012), Schaufeli et al. (2006), Villotti et al. (2014), Welfald et al. (2012), Willmer et al. (2019) and Yusoff et al. (2013), con los datos obtenidos en el presente estudio, la escala mostró una alta consistencia interna (α = 0.907).

***Job performance***

En el presente estudio se definió el job performance como los comportamientos, más que los resultados, de los miembros de una organización que son relevantes para los objetivos de la misma (Campbell, 1990). Para evaluar este constructo se empleó la traducción al español de la versión en inglés del Individual Work Performance Questionnaire (IWPQ-Versión 1.0) de (Koopmans et al., 2016), realizada a efectos de la presente investigación. El cuestionario tiene 18 proposiciones relacionadas con el comportamiento de las personas en el trabajo, y los participantes deben indicar con qué frecuencia presentaron cada una de las conductas en los últimos tres meses, en una escala Likert de seis intervalos (1 = nunca/casi nunca – 6 = casi siempre/siempre). Este cuestionario está compuesto por tres escalas que miden tres aspectos del constructo job performance, usados en el presente estudio como variables criterio, a fin de evaluar la validez predictiva de la UWES-9:

* Task performance, definido como las conductas que contribuyen, directa o indirectamente, a los aspectos técnicos centrales de la organización (Borman & Motowidlo, 1997); es decir, la competencia con la cual el individuo realiza las tareas centrales asociadas a su trabajo (Koopmans et al., 2011). Un mayor puntaje en esta sub-escala es indicativo de un mejor task performance.
* Contextual performance, definido como los comportamientos individuales que contribuyen al mantenimiento del ambiente social y psicológico en el que tiene lugar the task performance, y que van más allá de lo prescrito formalmente como objetivos del trabajo (Borman & Motowidlo, 1997; Koopmans et al., 2011; Rotundo & Sackett, 2002; Viswesvaran & Ones, 2000). Un mayor puntaje en esta sub-escala es indicativo de un mejor contextual performance.
* Counterproductive behaviors, que son los comportamientos que perjudican el bienestar de la organización (Koopmans et al., 2011; Rotundo & Sackett, 2002; Viswesvaran & Ones, 2000). Un mayor puntaje en esta sub-escala es indicativo de una mayor ocurrencia de comportamientos contraproducentes para la organización.

Se eligió este instrumento porque en su desarrollo se siguió un proceso muy riguroso, tanto en lo que respecta a la revisión de la literatura que sustentó la elección de los ítems, como en las distintas pruebas realizadas por sus autores originales (véase Koopmans et al., 2011, 2014, 2016; Koopmans, Bernaards, Hildebrandt, van Buuren et al., 2013; Koopmans, Bernaards, Hildebrandt, de Vet et al., 2013) y otros investigadores (véase Sebastián & Solana, 2016; Widyastuti & Hidayat, 2018) a fin de evaluar su comportamiento psicométrico. En este sentido, en distintos estudios se han hallado índices de consistencia interna de cada una de las escalas del IWPQ que varían entre 0.76-0.87 para task performance, 0.72-0.87 para contextual performance, y 0.76-0.89 para counterproductive behaviors (Koopmans et al., 2016; Ramos-Villagrasa et al., 2019; Gabini & Salessi, 2016; Widyastuti & Hidayat, 2018). Las distintas escalas de la adaptación al español del IWPQ-Versión 1.0 empleada en la presente investigación también mostraron una adecuada consistencia interna (Task performance: α = 0.733; Contextual performance: α = 0.793; Counterproductive behaviours: α = 0.771).

**Análisis de datos**

We conducted Confirmatory Factor Analysis (CFA) to fit different measurement models: (a) tres factores correlacionados (3F), (b) dos factores correlacionados (2F), y (c) un factor, ya sea unidimensional (1F) o esencialmente unidimensional.

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

Para la realización del análisis descriptivo de los ítems de la UWES-9 y los análisis de regresión dirigidos a evaluar la validez predictiva de la UWES-9 se uso the SPSS 25.0 program.

**Results**

**Preliminary Analysis**

En el caso del work engagement, las medias de los ítems variaron entre 5.07 y 5.41, con desviaciones típicas entre 0.815 y 1.113 (Table 1), indicando que, en términos generales los participantes reportaron altos niveles de work engagement. En este sentido, las distribuciones presentaron una asimetría negativa indicativa de que, en todos los ítems, los puntajes tendieron a agruparse hacia los valores altos de la escala, mostrando una homogeneidad moderada. Adicionalmente, la distribucion de los datos fue leptocúrtica para todos los ítems, excepto el ítem 2 (Table 1).

**CFA Model Fit**

Los modelos bifactor (classical bifactor and S-1 bifactor) no convergieron; el resto de modelos, como se muestra en la Tabla 3, mostraron una bondad de ajuste adecuada en los índices CFI, TLI y SRMR; pero, no en RMSEA y especialmente χ2 (ver Tabla 3 y Figura 1), que son muy sensibles a pequeños problemas de especificación correcta del modelo.

Los resultados del ANOVA evidenciaron que, comparando los tres modelos, el de tres factores ajusta significativamente mejor que el de dos, y el de dos ajusta significativamente mejor que el uno (Table 4).

Table 4.

*Resultados Obtenidos en el ANOVA Realizado para Determinar Diferencias en la Bondad de Ajuste de los Modelos 3F, 2F y 1F*

 La razón para que los tres modelos muestren un buen ajuste es la fuerte correlación que se observa entre los factores considerados (vigor [VIG]-dedicación [DED] = 0.926; vigor-absorción [ABS] = 0.916; dedicación-absorción = 0.946), lo que permite plantear como opción más parsimoniosa el modelo unidimensional con la ventaja asociada a este tipo de estructura, ya sea unidimensional o esencialmente unidimensional (i.e. bifactor), que es la de poder utilizar la puntuación global del test, en lugar de tener que aplicarlo como un instrumento multidimensional (Trizano-Hermosilla et al., 2021).

 Ahora bien, el hecho de que χ2 sea significativo, sin concurrir la circunstancia de un tamaño muestral grande, es indicativo de que existe cierto grado de misspecification. De hecho, Balducci et al. (2010), Klassen et al. (2012) and Lovakov et al. (2017) observaron correlación entre errores por solapamiento de contenido entre dos ítems de la dimensión Vigor: V1 (At my work, I feel bursting with energy) y V2 (At my job, I feel strong and vigorous). Además,

Klassen et al. (2012), en sus muestras de Australia y Canadá, and Lovakov et al. (2017) hallaron solapamiento de contenido entre dos ítems de la dimensión Absorption: A8 (I am immersed in my work) y A9 (I get carried away when I am working). Finalmente, Balducci et al. (2010) observaron correlación entre los errores de los ítems: V2 y D3 (I am enthusiastic about my job), y Lovakov et al. (2017) encontraron correlación entre los errores de los ítems: V5 (When I get up in the morning, I feel like going to work) y A6 (I am happy when I am absorbed in my work).

En estos casos, los análisis de los índices de modificación indicaban la perdida de ajute por la no inclusión de estas correlaciones.

 Del mismo modo que en las investigaciones de Balducci et al. (2010), Klassen et al. (2012) and Lovakov et al. (2017), en el presente estudio se constató la existencia de correlaciones entre los errores de los siguientes pares de ítems: V1 y V2, y A8 y A9; pudiendo afirmar así que existe solapamiento de contenido entre pares de ítems. Adicionalmente, nuestros resultados mostraron la existencia de una correlación entre errores significativa entre el item D7 de Dedication (I am proud of the work that I do) y el A8 de Absorption (I am immersed in my work). Incluyendo estas correcciones en los modelos (3F&CE, 2F&CE, 1F&CE) se observó que ahora la bondad de ajuste era excelente en todos los índices, incluyendo the χ2 test (ver Tabla 5).

Para determinar en términos de bondad de ajuste χ2 cuál era el mejor modelo se aplicó un ANOVA (ver Tabla 6) en el que no se observaron diferencias significativas entre los tres modelos considerados; por lo que, en caso de seguir el criterio de parsimonia, esto nos llevaría a elegir el modelo unidinensional como el más adecuado.

Table 6.

*ANOVA sobre las diferencias en la Bondad de Ajuste de los Modelos 3F&CE, 2F&CE y 1F&CE*

**Validez predictiva de la UWES-9**

Como consecuencia de que la estructura factorial del UWES-9 se ajusta a un modelo unidimensional, es legítimo utilizar la puntuación global del instrumento para la evaluar su validez predictiva. A su vez, como en el instrumento se distinguen tres dominios de contenido (vigor, dedicación y absorción), también se evaluó la capacidad predictiva de las sub-escalas.

Se utilizó como criterio el desempeño laboral autoevaluado, medido con la adaptación al castellano del IWPQ-Versión 1.0 y que informa de tres factores claves: task performance, contextual performance and counterproductive behaviour. Se realizaron análisis de regresión simple, utilizándose como predictor la puntuación global de la UWES-9.

Los resultados evidenciaron que, la puntuación global o WE del UWES-9 predice significativamente the task and the contextual performance; de forma que, en la medida en que los trabajadores muestran niveles más altos de WE, estos consideran que su desempeño es mejor, tanto en las tareas centrales asociadas a su trabajo como en aquellas que contribuyen al buen desempeño organizacional aun cuando no están formalmente asociadas a los objetivos del trabajo. Adicionalmente, se halló que the WE explicó un porcentaje más alto (33.4%) de la variabilidad del contextual performance, que de la varianza del task performance (23.8%). No obstante, the WE no predice significativamente the counterproductive behavior (Table 7).

Table 7.

*Resultados de las Regresiones Simples realizadas para Evaluar el Poder Predictivo del WE como Medida Unidimensional*.

Para determinar el poder predictivo de cada una de las dimensiones teóricas de la UWES-9 se realizó un análisis de regresión múltiple para cada uno de los aspectos del desempeño laboral empleados como variables criterio. Los resultados mostraron la relevancia de considerar las puntuaciones en cada una de las sub-escalas de la UWES-9, ya que ellas aportan distinta información predictiva que puede ser de utilidad dependiendo de cuál sea el propósito con el que se evalúe el WE.

En el caso del task performance, si bien el modelo predijo significativamente esta variable (R = 0.489; R2 = 0.239; F[1, 217] = 22.702, *p* = 0.000), la única dimensión de la UWES-9 que predijo significativamente the task performance fue el “vigor” (Table 8); de forma que, los participantes consideraron que su desempeño en las tareas directamente asociadas a su trabajo era superior, a medida que experimentaban mayores niveles de energía y resistencia física y mental mientras trabajan.

En cuanto al contextual performance, esta variable también fue predicha significativamente por el modelo (R = 0.584; R2 = 0.341; F[1, 217] = 33.446, *p* = 0.000), y tanto el “vigor” como la “absorción” predijeron significativamente the contextual performance, siendo el vigor el mejor predictor (Table 8). Así, las personas indicaron que ejecutaban con mayor frecuencia conductas que, aun sin estar explícitamente asociadas a los objetivos de su trabajo, contribuyen al adecuado desempeño de la organización, a medida que experimentaban mayor vigor y se sentían más concentradas y absortas en su trabajo.

Por último, en lo que respecta al counterproductive behaviours, esta variable fue predicha significativamente por el modelo (R = 0.199; R2 = 0.040; F[1, 215] = 2.965, *p* = 0.033); pero, la única dimensión de la UWES-9 relevante fue “dedicación” (Table 8); de forma que, en la medida en que las personas reportaban sentirse más entusiasmadas, inspiradas y orgullosas del trabajo que realizan, presentaban con menor frecuencia comportamientos perjudiciales para el bienestar de la organización. Lo hallado en cuanto al counterproductive behaviors permite explicar por qué cuando se emplea como predictor la puntuación en la escala global, ésta no predice significativamente the counterproductive behavior: la puntuación suma se anula porque the counterproductive behaviour se relaciona negativamente con el factor dedicación, lo que es esperable desde la teoría; pero, tiende a relacionarse positivamente con el factor absorción, lo que es contrario a lo esperable desde la teoría, ya que, en principio, lo que se esperaría es que la frecuencia de ocurrencia de conductas contraproducentes para la organización sea menor en los trabajadores que reportan sentirse más concentrados y absortos en su trabajo.

Table 8.

*Resultados de las Regresiones Múltiples realizadas para Evaluar el Poder Predictivo de las Dimensiones de WE.*

**Discussion**

El propósito de esta investigación fue determinar cuál es la estructura factorial de la UWES-9 más adecuada para la medida del constructo en Ecuador, comparando distintos modelos derivados de la teoría sobre work engagement (WE) y la evidencia empírica; así como, evaluar su validez predictiva respecto a la medida del desempeño laboral autoevaluado.

En lo que respecta a la estructura factorial de la UWES-9, en primer lugar, nuestros hallazgos no permiten apoyar las propuestas de autores como De Bruin and Henn (2013) and Fong and Ho (2015) respecto a la posible adecuación de un partial bifactorial model, pues the S-1 bifactor model con el factor dedicación como factor de referencia no covergió; and the classical bifactor model tampoco convergió, probablemente como consecuencia de la buena bondad de ajuste del modelo unidimensional. Seguramente en muestras en las que la correlación entre factores sea menor estos modelos que plantean unidimensionalidad esencial pueden ser una buena alternativa al modelo unidimensional.

En segundo lugar, los resultados de los CFA realizados para los modelo de tres factores correlacionados, dos factores correlacionados (dedicación+vigor, y absorción) and unifactorial, evidenciaron que los tres modelos mostraron índices CFI, TLI y SRMR indicadores de buen ajuste; pero, este no fue el caso de RMSEA y el χ2 test. Esto es consistente con lo reportado por Hallberg and Schaufeli (2006), Kataria et al. (2013), Klassen et al. (2012), Sousa-Vásquez et al. (2017), Villoti et al. (2014), Welfald et al. (2012) and Willmer et al. (2019). Ahora bien, en el presente estudio los modelos sí difirieron significativamente en términos de bondad de ajuste, de forma que el modelo 3F resultó significativamente superior a los otros dos. Esta superioridad del modelo de tres factores coincide con lo hallado por autores como Balducci et al. (2010), Breevaart et al. (2012), Gómez-Garbero et al. (2019), Lovakov et al. (2017), Rodríguez-Montalbán et al. (2014), Schaufeli et al. (2006), and Yusoff et al. (2013).

Sin embargo, este nivel de análisis obvia el hecho de que tanto en esta como en otras investigaciones se ha detectado un posible problema de mala especificación, pues la escala incluye algunos ítems cuyos contenidos se solapan. Concretamente, en la muestra ecuatoriana hemos confirmado los solapamientos de contenido observados por Balducci et al. (2010), Klassen et al. (2012) y Lovakov et al. (2017) entre los ítems V1 y V2, y por Klassen et al. (2012) y Lovakov et al. (2017) entre los ítems A8 y A9. Adicionalmente, en el presente estudio hemos detectado un nuevo solapamiento no reportado previamente entre los ítems D7 y A8 que debe investigarse en futuros estudios para determinar si es un problema específico de la muestra con la que se ha trabajado en la presente investigación, o si verdaderamente existe una relación entre estos dos indicadores que es generalizable a otras muestras y culturas.

 Cuando se realizan los CFA de forma de corregir este posible problema de redundancia, los tres modelos muestran excelentes índices de ajuste, no existiendo diferencias estadísticamente significativas entre ellos. En este punto nuestros resultados difieren de lo reportado por otros autores en el sentido de que, en nuestro caso, al permitir las correlaciones entre los errores, los modelos fueron equivalentes en cuanto a bondad de ajuste incluso considerando el χ2 test; mientras que, en todos los estudios consultados los autores concluyen sobre la superioridad o no de un modelo sobre otro sin considerar el valor de χ2 y en todas las investigaciones previas el χ2 fue estadísticamente significativo. De hecho, Fong and Ho (2015) señalan que uno de los problemas metodológicos que se encuentra con frecuencia en los estudios sobre el UWES-9 es que los investigadores ignoran los valores de χ2 significativos alegando que esta prueba es muy sensible para detectar desajustes del modelo cuando se trabaja con muestras grandes; pero, lo cierto es que el valor de este índice de ajuste suele ser ignorado incluso en estudios con muestras de tamaño pequeño-moderado.

Lo hallado en cuanto a la bondad de ajuste de los modelos al permitir las correlaciones entre los errores llevaría a pensar que, puesto que estructuras más complejas no suponen una ventaja sobre la más simple, por parsimonia se justificaría el uso la UWES-9 como instrumento unidimensional del WE. Esta unidimensionalidad presenta algunas ventajas desde el punto de vista métrico como lo es el poder utilizar la puntuación global de la prueba; lo cual, de acuerdo con Klassen et al. (2012) and Schaufeli et al. (2006), puede ser preferible para el uso de la UWES-9 across settings.

No obstante, esta opción colide claramente con la conceptualización teórica del WE como un constructo multidimensional; noción de la que partieron los creadores de la UWES-9 y quienes al desarrollar la escala no pretendieron construir un instrumento unidimensional, sino hacer una reducción en la cantidad de ítems que mantuviese la distinción entre las tres dimensiones. Se enfrenta así un conflicto entre el principio de parsimonia y la propuesta teórica, cuya aclaración requiere una breve reflexión sobre el modelo teórico y la bondad de ajuste de los modelos.

Es habitual confundir la unidimensionalidad con la homogeneidad de la medida del constructo. El hecho de que el modelo unidimensional muestre una bondad de ajuste equivalente a la de un modelo de tres factores, no implica que las dimensiones no existan, simplemente implica que desde el punto de vista de ajuste empírico no es posible diferenciar la respuesta a los ítems de cada dimensión. Así, nuestros resultados son una prueba de que, en la muestra analizada, la escala no permite diferenciar entre los factores. Esto solo indica que, en muestras diferentes, unos y otros componentes teóricos operan y se relacionan entre sí de distinta forma; en algunos, como fue nuestro caso, van tan parejos que no es posible diferenciarlos, en otras muestras pueden observarse dos de ellos o los tres elementos están tan diferenciados que el modelo más ajustado resulta el de dos factores o el de factores “no correlacionados”.

Ahora bien, las dimensiones existen desde el momento que el investigador ha enunciado ítems para la evaluación de estas dimensiones. Así, por ejemplo, el item 1 “En mi trabajo me siento lleno de energía” (Vigor), es un indicador que rastrea un dominio de contenido claramente diferente al item 7 “Estoy orgulloso del trabajo que hago” (Dedication), y estos a su vez difieren del ítem 8 “Estoy inmerso en mi trabajo” (Absorption). Otra cosa es que vigor, dedicación y absorción vayan de la mano, de modo que una persona que se sienta más involucrada e identificada con su trabajo, a su vez, le dedique mayor esfuerzo y se sienta más absorta cuando lo realiza.

En el ámbito educativo, en el que los profesionales suelen tener gran compromiso, no es extraño que las tres dimensiones estén fuertemente correlacionadas, y que el modelo más parsimonioso sea el unidimensional, lo que no debe confundirse con una medida simple o con la sub-representación del dominio de contenido. Como un constructo unidimensional, nuestros resultados muestran que, también en Ecuador, el WE predice significativamente the task and the contextual performance autoevaluado; lo cual se suma a la evidencia reportada por otros autores (Afacan-Findikli, 2015; Breevaart et al., 2015; Christian et al., 2011; Dalal et al., 2012; Gutermann et al., 2017; Kim et al., 2013; Lorente et al., 2014; Lupano et al., 2018; Motyka, 2018; Pongton & Suntrayuth, 2019; Qodariah et al., 2019; Rich et al., 2010). Además, se constató que the WE explicó un porcentaje más alto de la variabilidad del contextual performance, que de la varianza del task performance. Esto puede deberse a que, según Borman and Motowidlo (1997), una de las diferencias entre task and contextual performance es que es más probable que the contextual performance esté determinado por características relativamente estables de los individuos, como podría ser the WE; mientras que, los antecedentes of the task performance involucran más bien habilidades cognitivas. Ahora bien, a diferencia de lo que sería esperable, the work engagement no permitió predecir the counterproductive behaviors.

Lo anterior no significa que las dimensiones del WE sean irrelevantes; de hecho, lo son no solo desde el punto de vista teórico, sino porque ellas permiten predecir aspectos distintos del job performance. Concretamente, nuestros resultados mostraron que el task performance fue pedicho significativamente solo por el vigor, the counterproductive behaviours fueron predichas solo por la dedicación, y the contextual performance fue predicho tanto por el vigor como por la absorción, siendo el vigor el mejor predictor. Siguiendo estos resultados sobre la validez predictiva de la UWES-9, la recomendación sería emplear el modelo trifactorial, pues el modelo unidimensional lleva a la pérdida de información relevante.

**Conclusiones**

Partiendo de lo anteriormente expuesto podemos concluir que, en lo que respecta a la estructura de la UWES-9, nuestros resultados se suman a la evidencia previa disponible en cuanto a que es inapropiado generalizar automáticamente a cualquier contexto la noción del work engagement como un constructo conformado por las tres dimensiones planteadas por Schaufeli y sus colegas. En este sentido, se evidencia un problema en cuanto a la magnitud de las correlaciones entre los factores que se observan en distintas muestras y que se ve agravado por pequeños problemas de solapamiento de contenido entre algunos ítems, como revela las correlaciones entre errores observadas.

Así, debido a las consecuencias que se derivan de aceptar unos u otros modelos (e.g. si se puede usar la puntuación del total de la escala o bien trabajar como si fuera un instrumento multidimensional con las puntuaciones de cada sub-escala), coincidimos con autores como Klassen et al. (2012) al afirmar que el realizar en cada caso el CFA es clave para el investigador aplicado a la hora de entender bien cómo opera la medida del constructo en una determinada muestra y realizar los análisis pertinentes.

Las decisiones que se tomen en cuanto a la aceptación de una u otra estructura no deben estar guiadas solamente por la bondad de ajuste o el principio de parsimonia, la clave está en reconciliar la teoría con la evidencia empírica. En la búsqueda de esta reconciliación, al menos en el caso de Ecuador, aunque es posible utilizar la puntuación total ya que hay una estructura unidimensional aceptable, la propuesta de tres factores resulta superior desde un punto de vista sustantivo y predictivo.

**Limitaciones**

La principal limitación del presente trabajo es haber utilizado una simple muestra de una ocupación concreta y fundamentalmente de género femenino. Probablemente si se hubieran recogido múltiples muestras con características diferentes se habría podido reproducir toda la variabilidad de modelos propuestos en la literatura. Este ejercicio sería deseable para validar nuestra propuesta de que toda aplicación empírica debido a características propias de los participantes llevará a pequeñas distorsiones y que estas deben ser tenidas en cuenta para el uso correcto de las puntuaciones del instrumento.