

An Ultra-Short Measure of Positive and Negative Affect: The Reduced Affective Well-Being Scale (RAWS)

Pia H. Kampf^{1,*}, Ana Hernández²,
Vicente González-Romá³

¹ <http://orcid.org/0000-0003-1856-9716> / University of Valencia, Spain

² <http://orcid.org/0000-0002-5237-0535> / University of Valencia, Spain

³ <http://orcid.org/0000-0002-0657-7375> / University of Valencia, Spain

Abstract

In applied organizational research, where economy of scale is often a crucial factor in successful assessment, ultra-short measures are often needed. This study investigates the psychometric properties of the Reduced Affective Well-Being Scale (RAWS), an ultra-short measure of positive and negative affect in the workplace. This 6-item ultra-short version was compared with the original 12-item scale proposed by Segura and González-Romá (2003) in terms of internal consistency and criterion validity, using a sample of 1117 bank employees. In addition, longitudinal measurement invariance and within-subject reliability of the RAWS over time were assessed in a longitudinal sample of 458 employees at 12 time points. Results provide evidence that the RAWS is similar to the full scale in terms of reliability and validity. In addition, the RAWS shows satisfactory within-person reliability and factor loading invariance over time. In studies with intensive longitudinal designs that require repeated measures of affective well-being, the use of RAWS is a recommendable option.

Keywords: positive affect, negative affect, longitudinal factorial invariance, reliability, affect scale validation.

Uma Medida Ultracurta de Afeto Positivo e Negativo: A Escala Reduzida de Bem-Estar Afetivo (RAWS)

Resumo

Na pesquisa organizacional aplicada, onde a economia de escala é frequentemente um fator crucial para uma avaliação bem-sucedida, medidas ultracurtas são frequentemente necessárias. Este estudo investiga as propriedades psicométricas da Escala de Bem-Estar Afetivo Reduzido (RAWS), uma medida ultracurta de afeto positivo e negativo no local de trabalho. Esta versão ultracurta de 6 itens foi comparada com a escala original de 12 itens proposta por Segura e González-Romá (2003) em termos de consistência interna e validade de critério, usando uma amostra de 1117 funcionários do banco. Além disso, a invariância da medição longitudinal e a confiabilidade dentro do indivíduo do RAWS ao longo do tempo foram avaliadas em uma amostra longitudinal de 458 funcionários em 12 pontos no tempo. Os resultados fornecem evidências de que o RAWS é semelhante à escala completa em termos de confiabilidade e validade. Além disso, o RAWS mostra confiabilidade interna satisfatória e invariância de carga fatorial ao longo do tempo. Em estudos com desenhos longitudinais intensivos que requerem medidas repetidas de bem-estar afetivo, o uso de RAWS é uma opção recomendável.

Palavras-chave: afeto positivo, afeto negativo, invariância fatorial longitudinal, confiabilidade, validação de escala de afeto.

Una Medida Ultra Corta de Afecto Positivo y Negativo: La Escala Reducida de Bienestar Afetivo (RAWS)

Resumen

En la investigación organizacional aplicada, donde la economía de escala es a menudo un factor crucial para una evaluación exitosa, a menudo se necesitan medidas ultracortas. Este estudio investiga las propiedades psicométricas de la Escala de bienestar afectivo reducido (RAWS), una medida ultracorta de afecto positivo y negativo en el lugar de trabajo. Esta versión ultracorta de 6 ítems se comparó con la escala original de 12 ítems propuesta por Segura y González-Romá (2003) en términos de consistencia interna y validez de criterio, utilizando una muestra de 1117 empleados bancarios. Además, la invariación de medición longitudinal y la fiabilidad intraindividual del RAWS a lo largo del tiempo se evaluaron en una muestra longitudinal de 458 empleados en 12 puntos temporales. Los resultados proporcionan evidencia de que el RAWS es similar a la escala completa en términos de fiabilidad y validez. Además, el RAWS muestra una fiabilidad intrapersona satisfactoria e invariación de carga de factores a lo largo del tiempo. En estudios con diseños longitudinales intensivos que requieran medidas repetidas de bienestar afectivo, el uso de RAWS es una opción recomendable.

Palabras clave: afecto positivo, afecto negativo, invariación factorial longitudinal, fiabilidad, validación de escala de afecto.

Affective well-being – the frequent experience of positive affective responses such as joy or optimism and lack of negative affective responses such as sadness and tension (Luhmann et al., 2012; Eid & Larsen, 2008) – has gained importance in the context of work and organizations in the past few decades. Especially in highly developed countries, the focus on affective well-being in the workplace has been rapidly increasing as, with growing economic stability, priorities shift from purely financial concerns to general quality of life (Ilies et al., 2015). Individuals who report high affective well-being in their workplaces also tend to have better mental and physical health (Wilson et al., 2004), and they are happier in other areas of life and in life overall (Ilies et al., 2009; Sarwar et al., 2019). Therefore, it is desirable to foster affective well-being in employees, not only as a self-serving variable in the pursuit of individual happiness, but also from a business perspective. Higher affective well-being has been linked to better individual performance (Bryson et al., 2017; Drewery et al., 2016; Judge et al., 2001; Nielsen et al., 2017; Wright & Cropanzano, 2000), more organizational citizenship behaviors (Ilies et al. 2009), and less absenteeism (Lyubomirsky et al., 2005; Medina-Garrido et al. 2020) and turnover behavior (Wright & Bonnett, 2007).

Considering its high relevance in work-related contexts, affective well-being has been a central variable in many recent organizational studies. Understanding its fluctuations and dynamics over time and within individuals is crucial for increasing both the predictability and influenceability of its organizational outcomes. For this reason, as organizational researchers, we need an instrument that can reliably and validly assess change in AWB. Although there are a variety of instruments available to measure affective well-being in the workplace, an ultra-short measure with excellent psychometric properties is not yet available, and the 10-item International Positive and Negative Affect Schedule Short-Form I-PANAS-SF (Thompson, 2007; a short form of the Positive and Negative Affect Schedule; Watson et al., 1998) is currently the shortest validated measure.

Recruiting large samples in organizational settings often involves great effort and expense (Greiner & Stephanides, 2019). Therefore, researchers frequently assess a large number of variables at the same time in these studies, allowing them to test several research hypotheses. In this context, an economical questionnaire design is crucial because time and resources are often scarce, and it is difficult to achieve acceptable response rates (Baruch & Holtom, 2008). In addition, affective well-being has recently received special interest in studies with a dynamic focus that implement intensive longitudinal designs (ILDs) (see review on the dynamics of well-being by Sonnentag, 2015). ILDs involve taking many close, repeated measures across time from each surveyed subject. Examples of ILDs are event-sampling studies (e.g. Brosch & Binnewies, 2018) and diary studies (e.g. Meier et al., 2016). In these types of studies, participants have to respond to the same items repeatedly over a period of days or weeks, which can often be a nuisance for respondents – leading to the danger of low study adherence (Guo et al., 2016; Rolstad et al., 2011; Watson & Wooden, 2009). However, adherence is imperative to obtain high-quality data (Diehr et al., 2005): the more unresponsive the participants, the higher the probability of statistical bias (Groves, 2006; Tomaskovic-Devey et al., 1994). Therefore, we believe that there is a need for an ultra-short measure of affective well-being with excellent psychometric properties.

In the current study, we aim to assess the reliability and validity of an ultra-short measure of affective well-being used by Erdogan, Tomás, Valls, and Gracia (2018) and proposed in a research project carried out by our team (González-Romá & Hernández, 2013). This reduced affective well-being scale (RAWS)

is based on the 12-item affective well-being scale elaborated by Segura and González-Romá (2003), which has been successfully applied in organizational research in the past few years (Bashshur et al., 2011; Gamero, González-Romá, & Peiró, 2008; González-Romá & Hernández, 2016; González-Romá & Gamero, 2012). Segura and González-Romá's scale (2003) is composed of a set of six items that measure positive affect [three with a positive valence (cheerful, optimistic, lively) and three with a negative valence (sad, pessimistic, discouraged)] and a set of six items that measure negative affect [three with a negative valence (tense, jittery, anxious) and three with a positive valence (relaxed tranquil, calm)] . The RAWS is composed of the three positive-valence items that measure positive affect and the three negative-valence items that measure negative affect (below we explain the basis for this selection).

Specifically, we have three objectives. First, we aim to establish the RAWS' equivalence to its 12-item counterpart, examining both measures' inter-scale correlations and comparing their internal consistencies, as well as the sizes and patterns of relationships with relevant criterion variables in the organizational context. Second, by applying the RAWS in an intensive longitudinal design, we aim to estimate its within-subject reliability using multilevel confirmatory factor analysis to assess whether the scale reliably detects within-subject change over time (Geldhof et al., 2014). Third, we aim to assess the longitudinal measurement invariance of the RAWS, modelling its factor structure over time with four successively more restrained invariance models (Meredith, 1993): configural invariance (the same items load in the same factors over time), weak factorial invariance (in addition, item loadings are equal over time), strong factorial invariance (in addition, item intercepts are equal over time), and strict factorial invariance (in addition, item residuals are equal over time).

By achieving these objectives, we will contribute to the literature by providing a validation of the shortest measure – to the best of our knowledge – of affective well-being in the workplace to date. As a valid, reliable, and longitudinally robust ultra-short instrument, the RAWS can be a valuable asset in many types of organizational studies. Especially in ILDs, the RAWS could contribute to achieving stronger adherence rates and higher quality data over time. Researchers in the organizational context may consider it a worthwhile resource to improve the quality of the results in these types of studies.

Measuring Affective Well-being

Within and outside the organizational context, several models describe the construct of affective well-being (Segura & González-Romá, 2003). One of the most popular models that has received considerable empirical support over the years is the circumplex model of affect (Russell, 1980; see Figure 1). This model suggests that all affective states are defined along two independent bipolar dimensions: arousal and valence (Russell, 1980; Yik et al., 2011). Roughly, within these dimensions, there are four quadrants of affective qualities – for instance, affective qualities like joy are defined by high arousal and positive valence, qualities like tension are defined by high arousal and negative valence, qualities like depression are defined by low arousal and negative valence, and qualities like contentment are defined by low arousal and positive valence. Any affective quality can be defined in this two-dimensional space because all the possible types of affect surround these two dimensions in a circular fashion (George, 1990; Russell, 1980; Yik et al., 1999). Although some researchers have distinguished the four quadrants as distinct affective qualities (e.g. Burke et al., 1989; Mäkikangas et al., 2007), most research has focused on defining

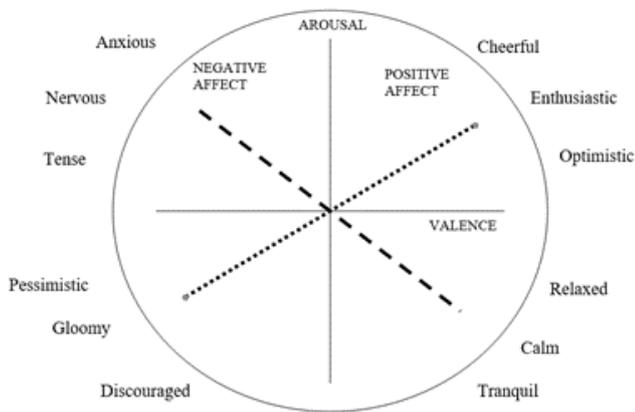


Figure 1. The circumplex model of affect (based on Russell, 1980).

two core affective dimensions, positive affect (PA) and negative affect (NA). Two different types of conceptualizations have been used to define these core dimensions in this model: either focusing on the orthogonal axes (i.e. PA summarizing all pleasant affects and NA summarizing all unpleasant affects) or focusing on the diagonal axes (i.e. PA describing pleasant activation vs. unpleasant deactivation and NA describing unpleasant activation vs. pleasant deactivation) (Warr et al., 2014). The latter conceptualization, with its focus on the diagonal rotation in the circumplex, has shown considerable discriminant validity in the relationships between PA and NA and key personality variables (e.g., the Big Five; Yik et al., 2002), behaviors (such as active voice and disengaged silence; Warr et al., 2014), and different neurophysiological systems (e.g., frontal lobe activation; Carver 2001; Carver & Scheier, 1991).

Affective well-being can be conceptualized as either a trait variable or state variable, with trait affect partly predicting state affective well-being (Kamarck et al., 2005; Nelis et al., 2016). Current organizational research, with its dynamic approach, is especially interested in affective well-being as a state, and in measuring within-subject change in this variable over time (Sonnentag, 2015). Levels of PA and NA rise and fall over time in the same person. These intra-individual fluctuations may depend on variables within the individual (such as personality: e.g. Larsen & Ketelaar, 1991) or outside the individual (such as job characteristics: Bakker, 2015). Understanding fluctuations in affective well-being and its dynamics over time and within individuals (rather than focusing solely on cross-sectional between-person differences) is crucial for increasing the predictability and influenceability of its organizational outcomes.

Popular scales to measure affective well-being in the workplace. In recent organizational studies, researchers have used several different measurement instruments to assess affective well-being. Many of these instruments have been validated across different contexts and situations and have good psychometric properties. There are three instruments that we want to mention specifically because they constitute the basis for most of the other scales that have been used to assess affective well-being in the workplace: Watson et al.'s (1988) positive and negative affect schedule (PANAS), Warr's affective well-being measure (1990), and Van Katwyck et al.'s (2000) job-related affective well-being scale (JAWS).

The PANAS is a 20-item instrument with two subscales, one for positive affect and one for negative affect, and it is rooted in the high-activation half of the affective circumplex. Each subscale consists of ten affective adjectives whose intensity is rated on a 5-point Likert-type scale ranging from "very slightly or not at all" to "extremely". The PANAS has demonstrated excellent internal consistency, stability over time, and convergent and discriminant

validity across general and specific samples (e.g. Carvalho et al., 2013; DePaoli & Sweeney, 2000; Von Humboldt et al., 2017; Watson et al., 1988). Almost two decades after the development of the original PANAS instrument, Thompson (2007) published a short form of the PANAS, the I-PANAS-SF, with only five items per subscale, thus shortening the original measure by half and providing a much more economical and time-efficient option to measure positive and negative affect. The I-PANAS-SF was deemed acceptable in terms of internal consistency, stability over time and across samples, and convergent and criterion validity (Thompson, 2007). Although the author of the scale encourages its application in different contexts, it was not designed to be context-specific to the organizational world, unlike the other scales presented here.

Warr's (1990) measure of affective well-being is based on the circumplex model of affect and focuses on the two diagonals in the circumplex. In the original scale, two bi-polar subscales are distinguished, depression-enthusiasm (positive affect) and anxiety-contentment (negative affect) (Warr, 1990), although some authors find a structure of four correlated yet distinguishable factors – depression, enthusiasm, anxiety, and contentment (Goncalves & Neve, 2011; Mäkikangas et al., 2007). The latter conceptualization has also been supported by the scale's author, with the specification that the scale can be used in different ways: to obtain an indicator for each quadrant in the circumplex, to obtain an indicator for positive and negative affect, respectively, or to obtain one general indicator for well-being, summarizing the whole circumplex (Goncalves & Neve, 2011; Warr & Parker, 2010). The scale consists of 12 affect items whose frequency is rated on a Likert-scale ranging from "never" to "all of the time". Because it was specifically conceived for job-related well-being, many researchers have used or adapted this instrument to measure affective well-being in the organizational context (e.g. Sevastos, 1996; Mäkikangas et al., 2007). However, although its reliability has been reported to be excellent (Warr, 1990; Sevastos et al., 1992), researchers have not always been able to confirm the underlying factor structure (Sevastos et al., 1992; Daniels et al., 1997). In order to further ensure reliability, increase support for the factor structure, and make the measure more applicable, Warr developed an extended 16-item scale, the IWP Multi-Affect Indicator, with four items for each quadrant of the affect circumplex (Warr & Parker, 2010; Warr, 2016).

Lastly, Van Katwyck and colleagues' (2000) JAWS is also a context-specific instrument designed to measure affective well-being in the workplace. It consists of 30 items that measure two subscales, positive and negative affect. Each affect valence subscale includes high-arousal items and low-arousal items. Items are scored for frequency on a 5-point Likert-scale ranging between "never" and "always". Some studies have applied a shortened adapted version of the JAWS, with six items in each subscale (Schaufeli & Van Rhenen, 2006; Van den Heuvel et al., 2015). The shortened version, like the original version, includes items from all the arousal levels for both positive and negative affect.

Although all these measures have significant merit in the organizational literature, they all present room for improvement in one crucial aspect for studies with an ILD: questionnaire economy. Some of the scales presented, such as the I-PANAS-SF with its 10-item structure, can be considered short scales that are quite economical. However, in many longitudinal studies, researchers assess a variety of different variables in the same sample because it can be quite difficult to recruit participants in applied organizational settings. In such cases, using reliable and valid scales with fewer items can help researchers to obtain high study adherence, good response rates over time, and high-quality

data (Biner & Kidd, 1994; Iglesias & Torgerson, 2000). Researchers implementing ILLDs that measure affect variables along with other hypothetical antecedents and consequences will benefit from a shorter scale with good psychometric properties.

The Reduced Affective Well-being Scale (RAWS)

The RAWS was proposed within a research project developed by our team (González-Romá & Hernández, 2013), and it was used by Erdogan et al. (2018) in a time-lagged study. It is an ultra-short measure of affective well-being based on the diagonal axes of the circumplex model of affect. As mentioned above, it is a reduced version of the affective well-being scale proposed by Segura and González-Romá (2003). The RAWS is composed of the three positive-valence items that measure positive affect (cheerful, optimistic, lively) and the three negative-valence items that measure negative affect (tense, jittery, anxious). The original reversed items from each subscale were removed for the following reasons. Although there is a popular belief among researchers that reversed items might reduce response bias, recent research has shown that reversed items may actually be counterproductive because they can confuse respondents, especially when high concentration is required (Suárez-Alvarez et al., 2018; Van Sonderen et al., 2013). This can result in unexpected results in the factor structure because an additional factor dependent on the wording of items – reversed vs. direct wording – can be introduced (González-Romá & Lloret, 1998).

Although Erdogan and colleagues (2018) reported excellent Cronbach's alphas for the RAWS, there has not been a systematic effort to evaluate and validate the psychometric properties of this scale. In addition, because we believe that an ultra-short measure of well-being is especially warranted for ILLDs, in this study we examine its measurement invariance over time and its within-subject reliability. This will show whether the RAWS' psychometric properties are invariant over time, and whether it can reliably detect within-subject changes in affective well-being over time.

Method

Participants, Data Collection Procedures and Ethical Considerations

We collected data from two samples for this study. Sample 1 was used for comparison analyses of the original 12-item scale and the reduced 6-item scale, whereas Sample 2 was used for longitudinal analyses of measurement invariance and within-subject reliability over time.

Sample 1. The data from Sample 1 were collected at two points in time between October and November 2012, and then again between October and November 2014. Data were collected at subsidiaries of a Uruguayan financial services provider, according to the company's possibilities and requests, using a pen-and-paper based self-report questionnaire. The questionnaires were administered by trained organizational psychologists with experience in applied organizational research. Research objectives were explained, and informed consent was given before each data collection session. The sample comprises 1117 participants, 37.2% of whom were women. Most of the employees were between 46 and 55 years old (68.1%); 20.9% were younger; and 11% were older. The majority of the respondents (69%) had already been working in this particular organization for more than 10 years at the time of the first data collection. With regard to education level, 53% of the respondents had a primary or secondary school

education level, 46% had some years of vocational training or university, and 1% had a graduate or postgraduate level university degree.

Sample 2. The data from Sample 2 were collected from October through December 2016. Informed consent was acquired from all participants. The project was approved by the ethics committee at the university. We hired a market research company to manage a respondent panel. This company invited the members of its Spanish panel to participate in the study, provided that they had a full-time employment relationship (not self-employed) and a secure work contract for at least three more months. Six hundred and twenty subjects agreed to participate. Subjects responded to an online questionnaire once a week for a period of 12 weeks (measurement occasions: T1-T12). Subjects who missed more than two measurement occasions in a row were not invited to participate again on further measurement occasions. In order to single out inattentive participants, we used three attention items (such as "Please choose response option 3 now") throughout the questionnaire at each measurement point. If a subject responded incorrectly to two or more of these items, their data for that measurement point were eliminated. A total sample of 458 subjects remained. Out of 5496 potential questionnaires (458 participants x 12 measurement points), we were provided with 3662 completed questionnaires, amounting to a response rate of 66.6%.

The sample provided a diverse mapping of the Spanish workforce. Respondents worked in companies from different sectors, ranging from industry and production via media and communication to research and education. Moreover, 50.9% were women, and the mean age was 39.6 years ($SD = 10.3$). Regarding education, 23.4% of respondents had secondary school education, whereas 65.6% had a basic university degree, and 11.1% had advanced university degrees. In terms of the hierarchy level within their companies, 3.7% worked in top management positions, whereas 51.3 had supervisory or middle management functions, and the remaining 45% were base level employees.

Instruments

The Affective Well-Being Scale. The Affective Well-Being Scale used by Segura and González-Romá (2003) was applied in Sample 1. It consists of two subscales (positive affect and negative affect) containing 6 items each, rated on a 5-point Likert scale ranging from 1 ("Not at all") to 5 ("Very much"). The positive affect subscale has three items with a positive valence ("To what extent did you feel happy / optimistic / lively this week at work") and three items with a negative valence ("To what extent did you feel sad / pessimistic / discouraged this week at work"). The negative affect subscale has three items with a negative valence ("To what extent did you feel tense / nervous / anxious this week at work") and three items with a positive valence ("To what extent did you feel relaxed / tranquil / calm this week at work") for negative affect. Previous studies using the scale have found that it has excellent reliability (e.g. $\alpha = .93$ for negative affect: González-Romá & Hernández, 2014; $\alpha = .92$ for positive affect: González-Romá & Gamero, 2012).

RAWS. The reduced version of Segura and González-Romá's (2003) scale, proposed by González-Romá and Hernández (2013) (the RAWS), has six items and was applied in Sample 2. The response scale is the same as the one used for the Affective Well-Being Scale. The positive affect subscale has three items with a positive valence ("To what extent did you feel happy / optimistic / lively this week at work" for positive affect). The negative affect subscale has three items with a negative valence ("To what extent

did you feel tense / nervous / anxious this week at work”). Erdogan et al. (2018) reported Cronbach’s alpha coefficients of .90 (positive affect) and .86 (negative affect). In Sample 1, we obtained the RAWs scores by selecting participants’ responses to the six items contained in the shortened version.

External criteria. We measured several correlates of affective wellbeing to compare the correlations between the RAWs and these correlates with those shown by the original Affective Wellbeing Scale (Segura & González-Romá, 2003).

Burnout. Burnout was measured in Sample 1 with a reduced version of the Maslach Burnout Inventory (Maslach et al., 1986). This reduced 9-item measure has previously been used to measure burnout in several organizational studies (Peiró et al., 2001; Schaufeli et al., 2005). The scale consists of three items for each of the three burnout subscales (cynicism, emotional exhaustion, and lack of personal accomplishment), measured on a 7-point response scale ranging from “never” to “every day”. Internal consistency was satisfactory for cynicism ($\alpha = .80$ for T1, $\alpha = .83$ for T2) and emotional exhaustion ($\alpha = .82$ for T1, $\alpha = .80$ for T2), and acceptable for lack of personal accomplishment ($\alpha = .69$ for T1, $\alpha = .69$ for T2).

Engagement. Engagement was measured in Sample 1 with the UWES-9 (Schaufeli et al., 2006). The scale consists of three items for each subscale (vigor, dedication, absorption) for engagement, rated on a 7-point response scale ranging from “never” to “every day”. Internal consistency was excellent for vigor ($\alpha = .88$ for T1, $\alpha = .90$ for T2), satisfactory for dedication ($\alpha = .80$ for T1, $\alpha = .79$ for T2), and acceptable for absorption ($\alpha = .70$ for T1, $\alpha = .66$ for T2).

Job performance. Job performance was measured with a 3-item scale (“How would you evaluate: 1. the overall quality of your work, 2. the quantity of work you produced, 3. your general job performance during the past week?”). Items 1 and 2 were extracted from Welbourne et al.’s (1998) role-based performance scale. Item 3 was based on Bond and Bunce’s (2001) one-item, self-rated performance scale. Items were rated on a 5-point response scale ranging between 1. “Very low” and 5. “Very high”. Internal consistency was excellent ($\alpha = .86$ for T1, $\alpha = .88$ for T2).

Data Analysis Procedures

In order to compare the 12-item scale with the RAWs 6-item scale, we performed correlation analyses with SPSS version 24. For all the remaining analyses, we used Mplus 8 (Muthén & Muthén, 2017), estimating model parameters via robust maximum likelihood. We first assessed configural invariance of the RAWs over time, by fitting, over the 12 time points, a two-factor model where the three items of PA and the three items of NA loaded, respectively, in two separate but correlated factors, PA and NA. Then, we sequentially tested for more demanding levels of longitudinal measurement invariance by imposing additional constraints: in addition to a consistent factor structure over time, item loadings were set to be equal over time (the weak-invariance model); in addition to equal factor loadings, the item intercepts were set to be equal over time (the strong-invariance model); and, in addition to equal factor loadings and intercepts, finally, item residual terms were set to be equal over time (the strict-invariance model) (see Meredith, 1993).

Finally, we performed a multilevel confirmatory factor analysis (MCFA; Muthén & Asparouhov, 2011) to estimate omega indicators of reliability (McNeish, 2018). MCFA is a specific multilevel structural equation model that is suitable for estimating reliability in multi-level data (Raykov & duToit, 2005). It makes it possible to separate the intra-individual portion from

the inter-individual portion of a factor model by allowing distinct estimations of between- and within-cluster covariance matrices (Geldhof et al., 2014). In the case of ILD data with different items that are responded to on several occasions, the intra-individual part of the model indicates factor loadings for the intra-individual variation (Bolger & Laurenceau, 2013). Intra- and inter-individual omegas as reliability indices can be obtained directly from the MCFA parameters (Bolger & Lorenceau, 2013). Because we are especially focusing on intra-individual change reliability, we aim to achieve a satisfactory intra-individual omega coefficient for both positive and negative affect.

For each model tested, we evaluated goodness of fit by means of the χ^2 -test, the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), the Tucker-Lewis-Index (TLI), and the Standardized Root Mean Square Residual (SRMR). For the RMSEA, values of $<.05$ indicate a good fit, and values around $<.08$ are acceptable (Chen et al., 2008; Hu & Bentler, 1999; Taasobshirazi & Wang, 2016). The CFI and TLI are considered acceptable when they exceed .90, and excellent when they exceed .95 (Hu & Bentler, 1999; Lai & Green, 2016; Taasobshirazi & Wang, 2016). In the case of the SRMR, a good fit is provided for values of $<.08$ (Hu & Bentler, 1999). For model fit comparison, we decided against using χ^2 difference tests. χ^2 -based statistics are extremely sensitive to sample size (Little, 1997), and so we based fit evaluations on alternative, more practical fit indices. Following Rutkowski and Svetina (2014), we compared relative fit based on the differences in CFI and RMSEA between the models. Typically, increases in CFI and decreases in RMSEA of $<.01$ are considered indicators of a similar model fit, which supports the invariance hypothesis from a practical point of view (e.g. Cheung & Rensvold, 2002; Putnick, & Bornstein, 2016). However, according to Rutkowski and Svetina’s (2014) simulations, when there are more than 10 groups to be compared (in our case more than 10 occasions, specifically, 12), the cut-off points for the differences in CFI and RMSEA should differ depending on the type of invariance (weak vs. strong) to be assessed. On the one hand, to conclude that metric or weak invariance holds, the decrease in CFI when comparing configural and weak invariance models should be less than .02, and the increase in RSMEA should be less than .03. On the other hand, for scalar or strong invariance, the standard cut-off point of .01 (e.g. Cheung & Rensvold, 2002) holds for both the CFI decrease and the RMSEA increase when comparing weak and strong invariance models. Considering that we had 12 measurement points, we followed Rutkowski and Svetina’s (2014) suggested cut-off points.

Results

Descriptive Statistics and Reliability

As a first step, we calculated the items’ means, standard deviations, and internal consistency for PA and NA, as assessed by the RAWs at both measurement occasions in Sample 1 (see Table 1) and for all twelve measurement occasions in Sample 2 (see Table 2).

In terms of reliability, the internal consistency of the RAWs was excellent for all occasions and samples, with a mean of $M\alpha = .91$ ($SD_{\alpha} = .01$) for NA and a mean of $M_{\alpha} = .93$ ($SD_{\alpha} = .01$) for PA in Sample 1, and a mean of $M\alpha = .93$ ($SD_{\alpha} = .01$) for NA and a mean of $M_{\alpha} = .94$ ($SD_{\alpha} = .02$) for PA in Sample 2. We also calculated the internal consistency for the original 6-item subscales in Sample 1, yielding a mean of $M_{\alpha} = .91$ ($SD_{\alpha} = .01$) for NA and a mean of $M\alpha = .91$ ($SD_{\alpha} = .02$) for PA. Internal consistencies for the 3- and 6-item scales in Sample 1 were similar, with a maximum

Table 1
Internal consistency, item means and standard deviations for negative affect and positive affect as measured by the RAW'S (Sample 1)

Measurement occasion	NA			PA				
	α	$M (SD)$			α	$M (SD)$		
		1 (Tense)	2 (Nervous)	3 (Anxious)		1 (Happy)	2 (Optimistic)	3 (Lively)
T1 (N = 1117)	.91	3.09 (1.12)	2.78 (1.20)	2.81 (1.21)	.94	3.04 (1.08)	2.98 (1.13)	3.06 (1.22)
T2 (N = 1181)	.90	2.76 (1.17)	2.49 (1.15)	2.57 (1.19)	.93	3.31 (1.04)	3.32 (1.06)	3.36 (1.06)

Note. NA = negative affect; PA = positive affect; T1-2 = measurement occasions 1 - 2

difference of .03 between Cronbach's alphas for any given 3-item subscale and its 6-item counterpart. Thus, we can conclude that the 3-item RAW'S scales show excellent reliability levels, similar to those of the original subscales.

in its three subscales (dedication, vigor, and absorption). We correlated all the affective wellbeing subscales at T1 with the criterion variables at T1 (cross-sectional correlations) and T2 (time-lagged correlations). Thus, we compared the two versions

Table 2
Internal consistency, item means and standard deviations for negative affect and positive affect, as measured by the RAW'S (Sample 2).

Measurement Occasion	NA			PA				
	α	$M (SD)$			α	$M (SD)$		
		1 (Tense)	2 (Nervous)	3 (Anxious)		1 (Happy)	2 (Optimistic)	3 (Lively)
T1 (N=458)	.89	2.67 (1.13)	2.66 (1.11)	2.542(1.14)	.91	3.22 (.92)	3.21 (.93)	3.26 (.89)
T2 (N=396)	.90	2.80 (1.08)	2.65 (1.07)	2.48 (1.12)	.93	3.12 (.93)	3.07 (.95)	3.08 (.95)
T3 (N=365)	.92	2.60 (1.13)	2.59 (1.07)	2.521(1.09)	.92	3.13 (.93)	3.12 (.97)	3.17 (.93)
T4 (N=354)	.90	2.66 (1.13)	2.62 (1.09)	2.50 (1.10)	.93	3.10 (.91)	3.06 (.93)	3.03 (.92)
T5 (N=325)	.92	2.60 (1.12)	2.61 (1.10)	2.49 (1.12)	.93	3.05 (.98)	3.02 (.96)	3.06 (.95)
T6 (N=285)	.93	2.44 (1.12)	2.46 (1.13)	2.38 (1.14)	.94	3.11 (.99)	3.06 (.95)	3.12 (.96)
T7 (N=277)	.93	2.41 (1.09)	2.38 (1.07)	2.33 (1.09)	.94	3.09 (1.00)	3.07 (1.03)	3.13 (1.03)
T8 (N=288)	.93	2.45 (1.08)	2.45 (1.09)	2.42 (1.12)	.95	3.09 (1.01)	3.06 (1.04)	3.10 (1.02)
T9 (N=236)	.93	2.31 (1.05)	2.28 (1.09)	2.29 (1.11)	.95	3.09 (1.04)	3.10 (1.04)	3.11 (1.05)
T10 (N=241)	.94	2.27 (1.10)	2.26 (1.09)	2.23 (1.08)	.95	3.13 (1.06)	3.12 (1.06)	3.19 (1.07)
T11 (N=244)	.93	2.39 (1.17)	2.40 (1.12)	2.33 (1.10)	.95	3.08 (1.06)	3.09 (1.08)	3.10 (1.07)
T12 (N=229)	.95	2.36 (1.12)	2.35 (1.14)	2.28 (1.13)	.97	3.14 (1.01)	3.13 (1.01)	3.12 (1.04)

Note. NA = negative affect; PA = positive affect; T1-12 = measurement occasions 1 - 12

Correlations between the 6-item original subscales and the 3-item subscales of the RAW'S

Using the data from Sample 1, we computed Pearson's bivariate correlation coefficients between the subscales of the original 12-item scale and the corresponding subscales of the RAW'S (see Table 3). Focusing on convergent validity between the 3-item and 6-item subscales, results show that correlations were .90 and .89, for positive and negative affective well-being ($p < .001$), respectively. Thus, the 3-item subscales seem to measure the same constructs (positive and negative affect) as their 6-item counterparts.

Table 3
Inter-scale Pearson's correlations between the original 6-item subscales and reduced 3-item subscales

	NA 3	PA 6	PA 3
NA 6	.90**	-.65**	-.60**
NA 3		-.60**	-.45**
PA 6			.89**

Note. NA 6 / PA 6 = original 6-item subscale for negative/positive affect; NA 3 / PA 3 = reduced 3-item subscale for negative/positive affect

Correlations with External Criteria

We further examined the correlations of the 6-item original subscales and the 3-item RAW'S subscales with different criterion variables: performance, burnout in its three subscales (exhaustion, cynicism, and lack of personal accomplishment), and engagement

of the PA and NA subscales in terms of "criterion validity" (concurrent and predictive validity). The correlations obtained are displayed in Table 4.

Correlation patterns were similar for the original and reduced scale versions, with differences between criterion correlations of the corresponding scales of between .00 and .10, and equal significance of the patterns of correlations, except in two cases: the correlations between the two versions of the NA scale and both cynicism and lack of personal achievement, although the correlation differences between the two scale versions were never larger than .06 (see Table 4). Additionally, we calculated Spearman's correlations (ρ) to test for the similarity of the correlation patterns shown by the RAW'S and Segura and González-Romá's (2003) subscales with the external criteria considered. The correlations obtained, $\rho = .99, p < .001$ for NA and $\rho = .999, p < .001$ for PA, showed that the pattern of correlations was similar for the two scale versions.

Longitudinal Invariance Analysis

We conducted this analysis using data from Sample 2. Model fit indices were excellent for the configural model and satisfactory for the weak, strong, and strict invariance models (see Table 5).

The CFI decrease when comparing configural and weak invariance models was .01, and the increase in RSMEA was .003. Thus, because they were below Rutkowski and Svetina's (2014) recommended cut-off points for metric or weak invariance (.02 and .03, respectively), the results show that the relationships

Table 4
Cross-sectional and time-lagged Pearson's correlations of the 6-item subscales and 3-item subscales with the criterion variables

Subscales	Performance		Burnout						Engagement					
			Exhaustion		Cynicism		Lack of Achievement		Vigor		Dedication		Absorption	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
NA 6	-.02	-.04	.57**	.37**	.16**	.15**	.22**	.16*	-.42**	-.28**	-.41**	-.31**	-.24**	-.19**
NA 3	.02	.00	.55**	.37**	.16**	.11	.14**	.08	-.34**	-.19**	-.31**	-.21**	-.15**	-.12*
PA 6	.16**	.11**	-.46**	-.28**	-.21**	-.22**	-.37**	-.14	.59**	.35**	.59**	.39**	.35**	.24**
PA 3	.19**	.13**	-.40**	-.21**	-.17**	-.21**	-.39**	-.14	.58**	.35**	.60**	.42**	.38**	.26**

Note. * $p < .05$, $p < .001$; two-tailed analysis. NA 6 / PA 6 = original 6-item subscale for negative/positive affect; NA 3 / PA 3 = reduced 3-item subscale for negative/positive affect. T1-2: measurement occasion 1 and 2 in Sample 1

between the items and the constructs are invariant over time (weak invariance is supported). The CFI decrease when comparing weak and strong invariance models was .014, and the increase in RSMEA was .004. In this case, the increase in RMSEA was below Rutkowski and Svetina's (2014) recommended cut-off points for scalar or strong invariance (.01). However, the increase in CFI exceeded the corresponding .01 cut-off point. The results of both indices are inconsistent. Thus, following a conservative approach, strong invariance was not supported by the data. Completely standardized factor loadings for the weak invariance model (the most parsimonious good-fitting model) ranged between .83 and .97, and they were all statistically significant ($p < .01$).

Table 5
Fit indices for factorial invariance of the RAW'S

Invariance models	χ^2 (df)	$\Delta\chi^2$	RMSEA	CFI	TLI	SRMR
configural	2919.27 (1812)**		.037	.957	.940	.032
weak	3234.99 (1858)**	315.72**	.040	.947	.927	.034
strong	3464.16 (1918)**	229.17**	.044	.933	.915	.074
strict	3575.79 (1924)**	111.63**	.044	.933	.912	.072

Note. ** $p < .01$. RMSEA: Root Mean Square Error of Approximation, CFI: Comparative Fit Index, TLI: Tucker-Lewis-Index; SRMR: Standardized Root Mean Square Residual

Multilevel Confirmatory Factor Analysis (MCFA)

The reliability of the RAW'S subscales to detect within-subject change over time in affective well-being was estimated by means of MCFA (Bolger & Lorenceau, 2013). Model fit indices for the two-factor MCFA model were excellent at both the between and within levels (see Table 6; $SRMR_w = .01$; $SRMR_b = .02$). The within-subject omega reliability coefficients (Ω) indicated satisfactory reliability of the RAW'S subscales for detecting intra-individual change over time, with $\Omega = .81$ for negative affect and $\Omega = .86$ for positive affect. In addition, between-subject omega coefficients indicated excellent reliability of the RAW'S subscales for detecting differences between individuals, with $\Omega = .98$ for negative affect and $\Omega = .96$ for positive affect.

Discussion

Our study's objectives were: 1) to establish the equivalence between the RAW'S and the full-length, 12-item affective well-

being scale proposed by Segura and González-Romá; 2) to provide evidence for the capacity of the RAW'S to reliably detect intra-individual change over time in both sub-dimensions of affective well-being, positive and negative affect; and 3) to assess whether the underlying factor structure proposed for the RAW'S holds and is replicable over time.

Regarding our first objective, which was to establish the correlations between the original 6-item subscales and the corresponding 3-item RAW'S subscales, the correlation coefficients obtained are not only statistically significant and high, but also close to 1, indicating that the long and shortened versions of the PA and NA subscales are measuring the construct in question in a similar way. With regard to internal consistency, Cronbach's alphas were excellent and similar for both scale versions. In addition, the correlation patterns of the 3-item RAW'S subscales with external criterion variables (performance, burnout, and engagement) were similar to the correlation patterns of the respective 6-item original subscales, for both the cross-sectional and time-lagged correlations. These results support the conceptual equivalence of the RAW'S subscales with their respective 6-item counterparts, indicating that the ultra-short scales may be used instead of the longer original versions without a significant loss of information and criterion-related validity.

Regarding our second objective, we found a satisfactory omega value in our MCFA for the within-person part of our model, indicating a satisfactory level of within-subject reliability of the RAW'S. This result supports the notion that the scale is suitable for reliably detecting intra-individual changes in affective well-being over time. The value obtained for omega-between shows that the RAW'S can reliably discriminate subjects' affective well-being.

Lastly, our third objective was to provide evidence for the longitudinal factor invariance of the RAW'S. Our results indicate satisfactory model fit for longitudinal measurement invariance across all the invariance models (configural, weak, strong, and strict). However, following a conservative approach, only weak invariance was supported because the differences between the weak invariance model and the strong invariance model were too large, according to one of the two relative fit indices used for model comparison. Thus, we can safely conclude that the relationships between item responses and latent factors are invariant over time. With regard to the RAW'S, we can say that the construct has the same meaning over time. Future research using new longitudinal data should clarify whether strong invariance is unambiguously supported, given that, in our study, one of the two

Table 6
Model fit and omegas for multilevel confirmatory factor analysis of the RAW'S

	χ^2	df	RMSEA	CFI	TLI	SRMR within	SRMR between	Ω NA within	Ω PA within	Ω NA between	Ω PA between
Model fit	40.24*	16	.021	.995	.989	.011	.018	.80	.84	.98	.96

Note. NA: negative affect, PA: positive affect. ** $p < .01$. RMSEA: Root Mean Square Error of Approximation, CFI: Comparative Fit Index, TLI: Tucker-Lewis-Index; SRMR: Standardized Root Mean Square Residual

indices considered supported this type of invariance. If this were the case, then it would be meaningful to compare the means over time (DeShon, 2004; Millsap & Heining, 2011). Hence, we can conclude that the RAWS is a suitable instrument for studies that investigate affective well-being over time.

Limitations and Strengths

As in every study, there are some limitations that researchers should take into account when interpreting the results reported here. First, all the study variables were measured by means of self-reports. Thus, some of the observed correlations may be inflated due to common-method variance. Future research seeking to provide further validity for the RAWS should consider measuring criterion variables that are not self-reported, such as objective performance or absenteeism. However, it is worth mentioning that most of the correlations shown by the RAWS were statistically significant (and showed the expected sign) when predicting the criteria two years after measuring positive and negative affect, with few exceptions. Thus, the results generally provide empirical evidence that supports the predictive power of the RAWS after two years, a time lag that somewhat mitigates concerns associated with common-method variance (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003).

Second, although establishing measurement invariance for the RAWS accomplishes a crucial step in ensuring the validity of affective well-being assessment in organizations, we want to stress that measurement invariance and within-subject reliability are not a property of a measure *per se*, but rather of the scores measured with the instrument. Therefore, we encourage researchers applying this scale in an IILD or other longitudinal designs to check the scale's properties in their own samples.

Our study also has some strengths. First, the IILD implemented in Sample 2 allowed us to test for measurement invariance and estimate within-subject reliability over time. The latter aspect is often forgotten in diary studies and investigations using experience-sampling methods. These studies usually try to ascertain whether within-subject variance in a predictor is related to within-subject variance in an outcome. Therefore, assessing whether the scales used can reliably detect within-subject changes over time is of the utmost importance (Bolger & Laurenceau, 2013). We urge researchers planning to develop IILDs to use multilevel CFA to estimate within-subject reliability. Second, we analyzed data from two different countries (Spain and Uruguay), and the RAWS showed good psychometric properties in both. These results offer some initial empirical evidence supporting the use of the RAWS in different countries.

Conclusion

In summary, this study provides valuable insight into the psychometric properties of the RAWS, and the results provide evidence for its equivalence with the original 12-item scale, within-subject reliability, and longitudinal metric or weak factorial invariance. In applied organizational research where economy of scale is often a crucial factor in successful assessment, ultra-short measures are often needed. Our results show that the use of this instrument in studies with longitudinal designs with repeated measures of affective well-being in organizational settings is an appropriate and valid option.

References

- Bakker, A. B. (2015). Towards a multilevel approach of employee well-being. *European Journal of Work and Organizational Psychology, 24*(6), 839-843. <https://doi.org/10.1080/1359432X.2015.1071423>
- Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations, 61*(8), 1139-1160. <https://doi.org/10.1177/0018726708094863>
- Bashshur, M. R., Hernández, A., & González-Romá, V. (2011). When managers and their teams disagree: a longitudinal look at the consequences of differences in perceptions of organizational support. *Journal of Applied Psychology, 96*(3), 558-573. <https://doi.org/10.1037/a0022675>
- Biner, P. M., & Kidd, H. J. (1994). The interactive effects of monetary incentive justification and questionnaire length on mail survey response rates. *Psychology & Marketing, 11*(5), 483-492. <https://doi.org/10.1002/mar.4220110505>
- Bolger, N., & Laurenceau, J. P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York, NY: Guilford Press.
- Brosch, E., & Binnewies, C. (2018). A diary study on predictors of the work-life interface: The role of time pressure, psychological climate and positive affective states. *MREV Management Review, 29*(1), 55-78. <https://doi.org/10.5771/0935-9915-2018-1-55>
- Burke, M. J., Brief, A. P., George, J. M., Roberson, L., & Webster, J. (1989). Measuring affect at work: confirmatory analyses of competing mood structures with conceptual linkage to cortical regulatory systems. *Journal of Personality and Social Psychology, 57*(6), 1091-1102. <https://doi.org/10.1037/0022-3514.57.6.1091>
- Carvalho, H. W. D., Andreoli, S. B., Lara, D. R., Patrick, C. J., Quintana, M. I., Bressan, R. A., ... & Jorge, M. R. (2013). Structural validity and reliability of the Positive and Negative Affect Schedule (PANAS): evidence from a large Brazilian community sample. *Brazilian Journal of Psychiatry, 35*(2), 169-172. <https://doi.org/10.1590/1516-4446-2012-0957>
- Carver, C. S. (2001). Affect and the functional bases of behavior: On the dimensional structure of affective experience. *Personality and Social Psychology Review, 5*(4), 345-356.
- Carver, C. S., & Scheier, M. F. (1991). *Self-regulation and the self. In The self: Interdisciplinary approaches* (pp. 168-207). New York, NY: Springer.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5
- Daniels, K., Brough, P., Guppy, A., Peters-Bean, K. M., & Weatherstone, L. (1997). A note on modification to Warr's measures of affective well-being at work. *Journal of Occupational and Organizational Psychology, 70*, 129-133.
- DePaoli, L. C., & Sweeney, D. C. (2000). Further validation of the positive and negative affect schedule. *Journal of Social Behavior and Personality, 15*(4), 561-568.
- Diehr, P., Chen, L., Patrick, D., Feng, Z., & Yasui, Y. (2005). Reliability, effect size, and responsiveness of health status measures in the design of randomized and cluster-randomized trials. *Contemporary Clinical Trials, 26*(1), 45-58. <https://doi.org/10.1016/j.cct.2004.11.014>
- Drewery, D., Pretti, T. J., & Barclay, S. (2016). Examining the Effects of Perceived Relevance and Work-Related Subjective Well-Being on Individual Performance for Co-Op Students. *Asia-Pacific Journal of Cooperative Education, 17*(2), 119-134.
- Eid, M., & Larsen, R. J. (Eds.). (2008). *The science of subjective well-being*. New York, NY: Guilford Press.
- Erdogan, B., Tomás, I., Valls, V., & Gracia, F. J. (2018). Perceived overqualification, relative deprivation, and person-centric outcomes: The moderating role of career centrality. *Journal of Vocational Behavior, 107*, 233-245. <https://doi.org/10.1016/j.jvb.2018.05.003>
- Gamero, N., González-Romá, V., & Peiró, J. M. (2008). The influence of intra-team conflict on work teams' affective climate: A longitudinal study. *Journal of Occupational and Organizational Psychology, 81*(1), 47-69. <https://doi.org/10.1348/096317907X180441>
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*(1), 72-91. <https://doi.org/10.1037/a0032138>
- George, J. M. (1990). Personality, affect, and behavior in groups. *Journal of Applied Psychology, 75*(2), 107-116. <https://doi.org/10.1037/0021-9010.75.2.107>
- González-Romá, V., & Gamero, N. (2012). Does positive team mood mediate the relationship between team climate and team performance? *Psicothema, 24*(1), 94-99.
- González-Romá, V., & Hernández, A. (2013). *Extending knowledge about graduates' overqualification in order to design intervention strategies*. Research Project funded by the Spanish Ministry of Economy and Competitiveness and the Spanish State Research Agency [Ref.: PSI2013-47195-R].

- González-Romá, V., & Hernández, A. (2014). Climate uniformity: Its influence on team communication quality, task conflict, and team performance. *Journal of Applied Psychology, 99*(6), 1042-1058. <https://doi.org/10.1037/a0037868>
- González-Romá, V., & Hernández, A. (2016). Uncovering the dark side of innovation: the influence of the number of innovations on work teams' satisfaction and performance. *European Journal of Work and Organizational Psychology, 25*(4), 570-582. <https://doi.org/10.1080/1359432X.2016.1181057>
- González-Romá, V. & Lloret, S. (1998). Construct validity of Rizzo et al.'s (1970) role conflict and ambiguity scales: A multisample study. *Applied Psychology, 47*(4), 535-545.
- González-Romá, V., Fortes-Ferreira, L., & Peiro, J. M. (2009). Team climate, climate strength and team performance. A longitudinal study. *Journal of Occupational and Organizational Psychology, 82*(3), 511-536. <https://doi.org/10.1080/026999498377737>
- Greiner, B., & Stephanides, M. (2019). Subject pools and recruitment. In A. Schram & A. Ule (Eds.), *Handbook of Research Methods and Applications in Experimental Economics* (pp. 313-334). UK: Edward Elgar Publishing. <https://doi.org/10.4337/9781788110563.00027>
- Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 646-675. <https://doi.org/10.1093/poq/nfl033>
- Guo, Y., Kopec, J. A., Cibere, J., Li, L. C., & Goldsmith, C. H. (2016). Population survey features and response rates: a randomized experiment. *American Journal of Public Health, 106*(8), 1422-1426. <https://doi.org/abs/10.2105/AJPH.2016.303198>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Iglesias, C., & Torgerson, D. (2000). Does length of questionnaire matter? A randomised trial of response rates to a mailed questionnaire. *Journal of Health Services Research & Policy, 5*(4), 219-221. <https://doi.org/10.1177/135581960000500406>
- Ilies, R., Aw, S. S., & Pluut, H. (2015). Intraindividual models of employee well-being: What have we learned and where do we go from here? *European Journal of Work and Organizational Psychology, 24*(6), 827-838. <https://doi.org/10.1080/1359432X.2015.1071422>
- Ilies, R., Fulmer, I. S., Spitzmuller, M., & Johnson, M. D. (2009). Personality and citizenship behavior: The mediating role of job satisfaction. *Journal of Applied Psychology, 94*(4), 945-959. <https://doi.org/10.1037/a0013329>
- Ilies, R., Wilson, K. S., & Wagner, D. T. (2009). The spillover of daily job satisfaction onto employees' family lives: The facilitating role of work-family integration. *Academy of Management Journal, 52*(1), 87-102. <https://doi.org/10.5465/amj.2009.36461938>
- Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict and performance in workgroups. *Administrative Science Quarterly, 44*(4), 741-763. <https://doi.org/10.2307/2667054>
- Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin, 127*(3), 376-407. <https://doi.org/10.1037/0033-2909.127.3.376>
- Kamarck, T. W., Schwartz, J. E., Shiffman, S., Muldoon, M. F., Sutton–Tyrrell, K., & Janicki, D. L. (2005). Psychosocial stress and cardiovascular risk: What is the role of daily experience? *Journal of Personality, 73*(6), 1749-1774. <https://doi.org/10.1111/j.0022-3506.2005.00365.x>
- Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research, 51*(2-3), 220-239. <https://doi.org/10.1080/00273171.2015.1134306>
- Larsen, R. J., & Ketelaar, T. (1991). Personality and susceptibility to positive and negative emotional states. *Journal of Personality and Social Psychology, 61*(1), 132-140. <https://doi.org/10.1037/0022-3514.61.1.132>
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*(1), 53-76. https://doi.org/10.1207/s15327906mbr3201_3
- Luhmann, M., Hofmann, W., Eid, M., & Lucas, R. E. (2012). Subjective well-being and adaptation to life events: a meta-analysis. *Journal of Personality and Social Psychology, 102*(3), 592-615. <https://doi.org/10.1037/a0025948>
- Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin, 131*(6), 803-855. <https://doi.org/10.1037/0033-2909.131.6.803>
- Mäkikangas, A., Feldt, T., & Kinnunen, U. (2007). Warr's scale of job-related affective well-being: A longitudinal examination of its structure and relationships with work characteristics. *Work & Stress, 21*(3), 197-219. <https://doi.org/10.1080/02678370701662151>
- Maslach, C., Jackson, S. E., Leiter, M. P., Schaufeli, W. B., & Schwab, R. L. (1986). *Maslach Burnout Inventory* (Vol. 21, pp. 3463-3464). Palo Alto, CA: Consulting psychologists' press.
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*, 412-433. <https://doi.org/10.1037/met0000144>
- Medina-Garrido, J. A., Biedma-Ferrer, J. M., & Sánchez-Ortiz, J. (2020). I Can't Go to Work Tomorrow! Work-Family Policies, Well-Being and Absenteeism. *Sustainability, 12*(14), 5519-5540. <https://doi.org/10.3390/su12145519>
- Meier, L. L., Cho, E., & Dumani, S. (2016). The effect of positive work reflection during leisure time on affective well-being: Results from three diary studies. *Journal of Organizational Behavior, 37*(2), 255-278. <https://doi.org/10.1002/job.2039>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525-543. <https://doi.org/10.1007/BF02294825>
- Muthén, B. O., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In *The handbook of advanced multilevel analysis*. J. Hox and J. M. Roberts (Eds.), 15–40. London: Taylor & Francis.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus: Statistical Analysis with Latent Variables: User's Guide* (Version 8). Los Angeles, CA: Muthén & Muthén
- Nelis, S., Bastin, M., Raes, F., Mezulis, A., & Bijttebier, P. (2016). Trait affectivity and response styles to positive affect: Negative affectivity relates to dampening and positive affectivity relates to enhancing. *Personality and Individual Differences, 96*, 148-154. <https://doi.org/10.1016/j.paid.2016.02.087>
- Nielsen, K., Nielsen, M. B., Ogbonnaya, C., Käsälä, M., Saari, E., & Isaksson, K. (2017). Workplace resources to improve both employee well-being and performance: A systematic review and meta-analysis. *Work & Stress, 31*(2), 101-120. <https://doi.org/10.1080/02678373.2017.1304463>
- Peiró, J.M., González-Romá, V., Tordera, N. & Mañas, M. A. (2001). Does role stress predict burnout over time among health care professionals? *Psychology and Health, 16*, 511-525. <https://doi.org/10.1080/08870440108405524>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71-90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Raykov, T., & Du Toit, S.H.T. (2005). Estimation of reliability for multiple component measuring instruments in hierarchical designs. *Structural Equation Modeling, 12*(4), 536-550. https://doi.org/10.1207/s15328007sem1204_2
- Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: is shorter better? A review and meta-analysis. *Value in Health, 14*(8), 1101-1108. <https://doi.org/10.1016/j.jval.2011.06.003>
- Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement, 74*(1), 31-57. <https://doi.org/10.1177/0013164413498257>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178. <https://doi.org/10.1037/h0077714>
- Sarwar, Bashir, & Khan (2019). Spillover of workplace bullying into Family Incivility: Testing a Mediated Moderation Model in a Time-Lagged Study. *Journal of Interpersonal Violence, 1-26*. <https://doi.org/10.1177/0886260519847778>
- Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement, 66*(4), 701-716. <https://doi.org/10.1177/0013164405282471>
- Schaufeli, W., González-Romá, V., Peiró, J. M., Geurts, S., & Tomás, I. (2005). Withdrawal and burnout in health care: On the mediating role of lack of reciprocity. In C. Korunka & P. Hoffmann. (Eds.), *Change and Quality in Human Service Work*, Volume 4 (pp. 205-226). München, Germany: Hampp Publishers.
- Schaufeli, W. B., & Van Rhenen, W. (2006). Over de rol van positieve en negatieve emoties bij het welbevinden van managers: Een studie met de Job-related Affective Well-being Scale (JAWS)[About the role of positive and negative emotions in managers' well-being: A study using the Job-related Affective Well-being Scale (JAWS)]. *Gedrag & Organisatie, 19*(4), 323-344.
- Segura, S. L., & González-Romá, V. (2003). How do respondents construe ambiguous response formats of affect items? *Journal of Personality and Social Psychology, 85*(5), 956-968. <https://doi.org/10.1037/0022-3514.85.5.956>
- Sevastos, P. (1996). *Job-related affective well-being and its relation to intrinsic job satisfaction* (Doctoral dissertation). Curtin University, Perth, Australia. Retrieved from <https://espace.curtin.edu.au/handle/20.500.11937/1909>
- Sevastos, P., Smith, L., & Cordery, J. L. (1992). Evidence on the reliability and construct validity of Warr's (1990) well-being and mental health measures. *Journal of Occupational and Organizational Psychology, 65*(1), 33-49. <https://doi.org/10.1111/j.2044-8325.1992.tb00482.x>

- Sonnentag, S. (2015). Dynamics of well-being. *Annual Review of Organizational Psychology and Organizational Behavior*, 2(1), 261-293. <https://doi.org/10.1146/annurev-orgpsych-032414-111347>
- Suárez-Álvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñoz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149-159. <https://doi.org/10.7334/psicothema2018.33>
- Suliman, A. M. (2001). Work performance: is it one thing or many things? The multidimensionality of performance in a Middle Eastern context. *International Journal of Human Resource Management*, 12(6), 1049-1061. <https://doi.org/10.1080/713769689>
- Taasobshirazi, G., & Wang, S. (2016). The performance of the SRMR, RMSEA, CFI, and TLI: An examination of sample size, path size, and degrees of freedom. *Journal of Applied Quantitative Methods*, 11(3), 31-40.
- Thompson, E. R. (2007). Development and validation of an internationally reliable short form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology*, 38(2), 227-242. <https://doi.org/10.1177/0022022106297301>
- Tomaskovic-Devey, D., Leiter, J., & Thompson, S. (1994). Organizational survey nonresponse. *Administrative Science Quarterly*, 39(3), 439-457. <https://doi.org/10.2307/2393298>
- Van den Heuvel, M., Demerouti, E., & Peeters, M. C. (2015). The job crafting intervention: Effects on job resources, self-efficacy, and affective well-being. *Journal of Occupational and Organizational Psychology*, 88(3), 511-532. <https://doi.org/10.1111/joop.12128>
- Van Katwyk, P. T., Fox, S., Spector, P. E., & Kelloway, E. K. (2000). Using the Job-Related Affective Well-Being Scale (JAWS) to investigate affective responses to work stressors. *Journal of Occupational Health Psychology*, 5(2), 219-230. <https://doi.org/10.1037/1076-8998.5.2.219>
- Van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PloS One*, 8(7), e68967-e68967. <https://doi.org/10.1371/journal.pone.0068967>
- Von Humboldt, S., Monteiro, A., & Leal, I. (2017). Validation of the PANAS: A measure of positive and negative affect for use with cross-national older adults. *European Studies Review*, 9(2), 10-19. <https://doi.org/10.5539/res.v9n2p10>
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070.
- Watson, N., & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. *Methodology of Longitudinal Surveys*, 1, 157-182.
- Warr, P. (1990). The measurement of well-being and other aspects of mental health. *Journal of Occupational Psychology*, 63(3), 193-210. <https://doi.org/10.1111/j.2044-8325.1990.tb00521.x>
- Warr, P., Bindl, U. K., Parker, S. K., & Inceoglu, I. (2014). Four-quadrant investigation of job-related affects and behaviours. *European Journal of Work and Organizational Psychology*, 23(3), 342-363. <https://doi.org/10.1080/1359432X.2012.744449>
- Warr, P., & Parker, S. (2010). *IWP multi-affect Indicator*. Unpublished instrument. Sheffield, UK: Institute of Work Psychology, University of Sheffield.
- Welbourne, T. M., Johnson, D. E., & Erez, A. (1998). The role-based performance scale: Validity analysis of a theory-based measure. *Academy of Management Journal*, 41(5), 540-555. <https://doi.org/10.5465/256941>
- Wilson, M. G., Dejoy, D. M., Vandenberg, R. J., Richardson, H. A., & McGrath, A. L. (2004). Work characteristics and employee health and well-being: Test of a model of healthy work organization. *Journal of Occupational and Organizational Psychology*, 77, 565-588. <https://doi.org/10.1348/0963179042596522>
- Wright, T. A., & Bonnett, D. G. (2007). Job satisfaction and psychological well-being as nonadditive predictors of workplace turnover. *Journal of Management*, 33, 141-160. <https://doi.org/10.1177/0149206306297582>
- Wright, T. A., & Cropanzano, R. (2000). Psychological well-being and job satisfaction as predictors of job performance. *Journal of Occupational Health Psychology*, 5(1), 84-94.
- Yik, M. S., Russell, J. A., Ahn, C. K., Dols, J. M. F., & Suzuki, N. (2002). Relating the five-factor model of personality to a circumplex model of affect. In *The five-factor model of personality across cultures* (pp. 79-104). Boston, MA: Springer.
- Yik, M. S., Russell, J. A., & Barrett, L. F. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology*, 77(3), 600-619. <https://doi.org/10.1037/0022-3514.77.3.600>
- Yik, M., Russell, J. A., & Steiger, J. H. (2011). A 12-point circumplex structure of core affect. *Emotion*, 11(4), 705-731. <https://doi.org/10.1037/a0023980>

Information about corresponding author:

Pia Helen Kampf

University of Valencia, Faculty of Psychology, IDOCAL, Avda. Blasco Ibañez 21, 46010 Valencia, Spain.
E-mail: pia.kampf@valencia.edu

Author note:

This study was supported with a research excellence grant from the Generalitat Valenciana with the reference number PROMETEO/2016/138.