



Advancing Research and Practice Through an Empirically Validated Short-Form Measure of Work Engagement

Russell A. Matthews¹  · Maura J. Mills¹ · Shelby Wise²

Received: 12 March 2020 / Revised: 4 June 2020 / Accepted: 16 July 2020 /
Published online: 31 July 2020
© Springer Nature Switzerland AG 2020

Abstract

Taking a multifaceted approach toward establishing validity, we present evidence from four independent samples supporting the efficacy of a short-form (three-item) version of the Utrecht Work Engagement Scale. The measure provides a psychometrically sound instrument to facilitate both complex and longitudinal research designs on engagement, as well as in practice, where parsimony is vital. To this end, we examine the short-form measure for measurement invariance based on a large heterogeneous sample (Study 1 Sample 3), as well as temporal invariance based on six waves of data with two-week lags between assessments (Study 2). As such, to demonstrate the utility of the short measure, we provide a within-individual test of the concept of gain spirals within conservation of resources theory (Study 2) wherein we introduce the notion top-down versus bottom-up gain spirals. Collectively we present strong validity evidence for the short-form measure such that it is applicable for both practitioners and academics alike.

Keywords Engagement · Short-form · Validation · Invariance · Life satisfaction

Arguably “the single most important issue in management research and practice in the last decade” (Saks 2017, p. 76), *work engagement*, has received considerable and increasing attention (e.g., Saks and Gruman 2014; Seppälä et al. 2015) in the realms of both research and practice. However, engagement – defined as a “positive, fulfilling,

✉ Russell A. Matthews
ramatthews2@ua.edu

Maura J. Mills
mjmill@culverhouse.ua.edu

¹ Department of Management, The University of Alabama, 361 Stadium Drive, Box 870225, Tuscaloosa, AL 35487, USA

² Department of Psychology, Bowling Green State University, Bowling Green, OH 43403, USA

and work-related state of mind characterized by vigor, dedication and absorption” (Schaufeli and Bakker 2004, p. 295) – has also seen widespread decline amongst employees worldwide. Gallup estimates that these disengaged employees are costing US companies alone upwards of \$250 billion annually (Rath and Conchie 2009). With these underling financials as a salient backdrop, focus on this construct also stems from its established relationship with a host of outcomes essential for organizational success, including employee productivity, firm profitability, and competitive advantage (e.g., Buckingham and Coffman 1999; Harter et al. 2002; Rich et al. 2010; Schaufeli 2013).

Nevertheless, despite this sustained importance and relevance, there remains a notable research-practitioner gap in terms of operationalizing the construct, as well as its effective measurement (Bailey 2016; Saks 2017). Indeed, scholars (e.g., Macey and Schneider 2008; Mills et al. 2013; Saks 2017; Shuck et al. 2017) have consistently lamented that, although engagement is relatively easy to recognize, it has been difficult to define and consistently measure – an issue that Saks (2017) calls “the engagement measurement barrier” (p. 77). This point was illustrated by Macey and Schneider (2008) who provided a synthesis of the many approaches, definitions, and expected behaviors associated with engagement in both practice and academia. The overall evaluation was that, with the exception of only a few (Salanova et al. 2005; Schaufeli et al. 2002), the majority of engagement measures have fallen short in truly representing the construct. Moreover, even of those few that do effectively represent engagement, most are generally too lengthy to be usable by practitioners or by researchers pursuing longitudinal designs. Such exclusions inherently limit those measures’ relevance for the most rigorous and impactful measurement efforts.

In an effort to provide clarity surrounding the measurement and operationalization of engagement, academics have engaged in considerable effort to validate measures of engagement in a transparent way (Christian et al. 2011; Mills et al. 2012; Wefald et al. 2012). Although a number of engagement measures exist (see Saks 2017), academia has seen widespread adoption of the conceptualization put forth by Schaufeli and Bakker (2004) and the associated Utrecht Work Engagement Scale (UWES; Schaufeli et al. 2002). As the most commonly used measure in academic research (Rich et al. 2010; Schaufeli 2012), the UWES has been found to relate to many outcomes important to employees and organizations (see Halbesleben 2010 for an overview). Practitioners, however, vary widely in their operationalization and measurement of engagement (Bakker and Leiter 2010; Macey and Schneider 2008) to the extent that it is questionable whether they are even assessing the same construct.

Moreover, with so many practitioners (e.g., consulting firms, in-house organizational behavior specialists) examining engagement, those in industry often find themselves further divorced *from one another* in terms of the operationalization and assessment of engagement (Mills et al. 2013). Although practitioners argue their instruments tap engagement and predict critical outcomes, their items are often proprietary, making it difficult to understand what exactly is being measured or how their assessments align with academic approaches. This Tower of Babel-type effect between, and within, practitioners and academics significantly hampers progress in the realms of both research and application. As emphasized by Shuck et al. (2017), “numerous entangled definitions, words, measurements, and frameworks have been proposed when referring to employee engagement... Consequently, researchers have routinely drawn theoretical

conclusions about the meaning of employee engagement, limiting [its] applicability in theory building and practice” (p. 263).

To systematically advance work engagement scholarship, the science-practitioner gap must be addressed. We argue that a ready avenue for doing so is via the validation of a parsimonious, psychometrically sound measure appropriate for use in practice as well as in rigorous academically-based research designs. Without such a measure, practitioners are limited in the extent to which they can leverage empirical research, while simultaneously using measures of engagement that are misaligned with the dominant, psychometrically-supported operationalization. Notably, only about half of organizations report employing analytic techniques at all, with even fewer using predictive methodology or more complex modeling (Collins et al. 2017). Organizations can effectively leverage empirical findings from advanced statistical methods, but only if their measurement and operationalization of the construct aligns with that used in the more rigorous academic research. But, that will not happen as long as the existing validated measures in academia are impractically lengthy. A psychometrically valid parsimonious measure is needed to help bridge the divide, thereby better positioning academics to collaborate with practitioners regarding a construct of practical interest to organizations (Bailey 2016). Put simply, a validated measure would allow academics and practitioners to speak a common language when examining issues related to employee engagement and how to best promote it throughout organizations.

A major hurdle facing practitioners – and academics collaborating with organizations – when trying to employ existing scholarly measures is that they are too lengthy to be accepted by organizational stakeholders and/or clients (Lapierre et al. 2018). Moreover, short-form measures, when they exist, are often compromised by nonexistent or limited validity evidence (Fisher et al. 2016; Smith et al. 2000) and even often fail to explain the decision rules used to adapt (e.g., shorten) a given measure (Heggestad et al. 2019). Especially in the case of lengthy annual surveys frequently used within organizations, the goal is usually to measure many constructs simultaneously, making lengthy measures for any one construct largely impossible from the perspective of practitioners and organizations. With these issues in mind, within this line of research we develop and thoroughly evaluate a short-form version of the UWES and its psychometric characteristics, giving due attention to clarifying the decision rules used to adapt the measure (Heggestad et al. 2019).

As we will discuss, some effort has been made in this respect (Reina-Tamayo et al. 2017; Schaufeli et al. 2017). However, existing validity evidence for these measures is limited. As such, within the current program of research, we aim to provide evidence regarding the utility of a short-form measure based on a multifaceted validity approach. In so doing, we make several contributions to the literature. Specifically, we offer a short-form measure aligning with Schaufeli et al.’s (2002) conceptualization of engagement, take a multifaceted validation approach to present extensive evidence for the efficacy of this measure across two studies and four data sources, evidence the temporal invariance (factorial/configural, metric, scalar, strict) of engagement as measured via the short-form measure, assess the extent to which engagement as measured is invariant across various demographic characteristics, and evaluate engagement’s over-time relationship with employee life satisfaction as an example of the utility of the measure for longitudinal research. In so doing, we also move toward closing the research-practitioner gap in the operationalization and measurement of engagement, thereby

facilitating implementation in both research and practice. In sum, regardless of application context, the collective evidence presented here should allow individuals (academic or practitioner) to use the resulting measure with the highest level of confidence regarding the underlying psychometrics of the measure.

Work Engagement Matters. But Why?

Several systematic reviews (e.g., Christian et al. 2011; Halbesleben 2010) elaborate on the criticality of work engagement. In brief, work engagement consistently relates to a myriad of employee and organizational outcomes, including decreased absenteeism (Bakker and Schaufeli 2008), turnover (Halbesleben 2010), decreased counterproductive work behaviors and increased organizational citizenship behaviors (Sulea et al. 2012), as well as employee retention (Buckingham and Coffman 1999). Engagement is also related to those bottom-line outcomes of importance to the success of organizations, such as in-role performance broadly defined (Christian et al. 2011; Halbesleben and Wheeler 2008), sales (Xanthopoulou et al. 2009) and service performance (Chen et al. 2018; Salanova et al. 2005), and firm profitability (e.g. Harter et al. 2002; Rich et al. 2010; Schaufeli 2013). Because work engagement is related to so many desirable organizational outcomes, it has been widely utilized in the applied arena and is a frequent construct of interest among consultancies' clientele – and this interest shows no signs of abating (Collins et al. 2017).

However, as noted, academics' and practitioners' respective understandings of engagement are often not well aligned (Bakker and Leiter 2010; Harter et al. 2002). Overwhelmingly, however, large firms and organizations claim to have evidence indicating that their measures relate to those same outcomes that academics have examined (Schaufeli and Bakker 2010). Although this may be true in terms of general direction of relations, it is unlikely that academic and practice-based measures of engagement are truly measuring the same thing. Rather, they are likely tapping into related, but different, underlying construct spaces. With this in mind, it is important to recognize that the relationships explicated above were identified using measures developed within academic engagement paradigms (with one exception; Gallup's Q12 – see Harter et al. 2002).

The Measurement of Work Engagement

Researchers (e.g. Bakker et al. 2011; Rich et al. 2010) have consistently emphasized the lack of clarity surrounding the measurement of engagement. Three primary studies (Byrne et al. 2016; Viljevac et al. 2012; Wefald et al. 2012) have been conducted to compare the psychometric and predictive efficacy of existing engagement measures (i.e., Britt et al. 2013; May et al. 2004; Schaufeli and Bakker 2004; Shirom 2003). Results suggest that the UWES is a superior measure as it better differentiates engagement from similar constructs (Viljevac et al. 2012) and is most predictive of relevant outcomes (Byrne et al. 2016; Wefald et al. 2012). Byrne et al. further argued that, in practice, the UWES conceptualization of engagement is likely most appropriate because it assesses a broader domain and yields practical information actionable within

organizational contexts. Although outstanding measurement issues remain (May et al. 2004; Saks and Gruman 2014; Shuck et al. 2017; Wefald et al. 2012), the UWES demonstrates strong psychometric characteristics, which likely explains why it remains so widely used by researchers. As such, we focus our efforts on improving the utility of this measure.

In line with this conceptualization, scholars characterize engagement predominantly as an affective-cognitive state comprised of vigor, dedication, and absorption (Schaufeli et al. 2002). These three engagement components align with Kahn's (1990) conceptualization of engagement. *Vigor* aligns with Kahn's physical-energetic component and is defined as liveliness and energy at work and encompasses ones' willingness to invest time in others and be mentally resilient in the face of difficulties. *Dedication* describes the extent to which an employee is involved, committed, and prideful about their work, and corresponds with Kahn's emotional component. *Absorption* is the extent to which an employee feels engrossed in his or her work and senses time as passing quickly, thus aligning with Kahn's cognitive component. With this operationalization in mind, Schaufeli et al. (2002) created the UWES, consisting of 17 items nested within the three conceptual factors. In brief though, research has been mixed regarding the factor structure of the measure (e.g., Sonnentag 2003; Mills et al. 2012).

A 9-item version of the UWES attempted to address these issues as well as to provide a more parsimonious measure. This version yielded correlations of $>.90$ for the three latent factors (Schaufeli et al. 2006), leading Schaufeli et al. to posit that this version might be best interpreted as holistic engagement rather than as three components, as the multifactor structure was inconclusive for the given sample. This notion of scoring the UWES-9 as a single unitary engagement index is, in part, reflective of the fact that conceptual factors often correlate with one another highly enough so as to indicate possible multicollinearity (e.g., .88, .94; Balducci et al. 2010; Schaufeli 2006). Beyond that, scholars (e.g., Balducci et al. 2010; De Bruin and Henn 2013) have recommended that the total score may be best practice for interpreting the UWES-9, as independently the three constructs may lack discriminant validity.

Current Program of Research

To date, two known studies have published adapted three-item versions of the UWES. While not their primary focus, Reina-Tamayo et al. (2017) selected three items to develop an "episodic" short-form measure (one item from each dimension of the UWES) for use in their diary study. Reina-Tamayo et al. report retaining the item from each dimension with the highest item-total correlation from the multi-item scale. Alternatively, Schaufeli et al. (2017, p. 4) offered a three-item short-form measure wherein they report selecting items "[b]ased on face validity, theoretical reasoning, and earlier feedback from respondents". These past efforts provide instructive "proof of concept" support for our argument regarding the need for a short engagement measure. However, while both studies contribute to our collective understanding of this need, there remain important areas of opportunity to improve on and more thoroughly validate a short-form engagement measure.

Specifically, a more methodological (objective) process for item selection as well as a more thorough consideration of validity evidence, based on established best practices

(e.g., Stanton et al. 2002) would ultimately allow for greater confidence in the final items in the resulting measure (Heggestad et al. 2019). As noted by Heggestad et al. (2019), it is common practice for researchers to adapt measures to fit their purposes (e.g., shorten a measure by selecting certain items to be retained), albeit with limited attention to the degree to which the adapted measure remains content valid. To that end, we conducted two separate studies, utilizing four unique data sources, to validate a proposed short-form of the UWES, with particular attention given to the necessary conditions for optimal construct measurement and assessment (e.g., Ployhart and Vandenberg 2010). As such, we address issues noted by Heggestad et al. (2019). That is, we provide scholars and practitioners a common extensively validated short-form measure to work from; that is, if applied consistently, utilizing the validated measure from this program of research would ensure that observed results across studies (or between research and practice domains) do not differ as a function of selecting a different subset of items.

Given that the UWES often yields near-multicollinear correlations between factors and therefore is best indexed in terms of collapsing across components (Christian et al. 2011; Sonnentag 2003), we used a multifaceted approach to validate a short-form measure that leveraged the strongest item, defined and clearly articulated here based on multiple criteria, from each of the conceptual factors. In Study 1 we determined the most representative item from each factor in line with recommendations by Stanton et al. (2002). Specifically, judgmental qualities of all items were evaluated by subject matter experts (SMEs; sample 1); that is, rather than relying on the potentially idiosyncratic judgments of the authors (Heggestad et al. 2019), we developed a holistic understanding based on larger number of individuals actively involved in the engagement research domain. Following this, internal and external qualities of the items were evaluated to determine factor loadings and the relationship to related constructs (samples 2, 3). In Study 1 we also examined the extent to which this engagement measure functions comparably across varying demographic characteristics. That is, with practitioners and researchers alike increasingly seeking to compare demographic groups on engagement, it is important to first establish comparable measurement functioning across groups to ensure unbiased comparison.

Subsequently in Study 2, we first examined the psychometric characteristics of the short-form measure (i.e., validity evidence for the short-form measure independent of the administration of the long-form; Smith et al. 2000). In so doing, we expand our understanding of the validity of this measure by examining it for temporal factor invariance (6 waves of data with two-week lags between assessments). In turn, to demonstrate the utility of the short measure (and establish predictive validity evidence), we provide a within-individual test of the concept of gain spirals within conservation of resources theory (Study 2) and examine the dynamic (over time) cause-effect relationship between engagement and life satisfaction.

Study 1

In Study 1 we used data from three independent samples to determine, as objectively as possible, the best items for the short-form measure following best practice recommendations (e.g., Stanton et al. 2002; Smith et al. 2000). Sample 1 consisted of SMEs

familiar with the engagement construct and its measurement. To examine issues related to *judgmental* qualities, SMEs were asked to evaluate engagement items from the UWES-9 (Schaufeli et al. 2006) for conceptual fit with their respective dimension (i.e., vigor, dedication, absorption). Samples 2 (working adults collected via Mechanical Turk) and 3 (working adults collected through a peer-nomination approach) were used to examine internal and external qualities of items from the UWES-9 to help identify the final set of items and to further assess the validity of the abbreviated measure. In turn, we applied a stressor taxonomy approach (i.e., Halbesleben 2010) to drive our framing and understanding the internal and external qualities of items from the UWES. This approach was taken given existing research has and continues to consistently link engagement to various stressors [e.g., role stressors & work hours (Kronenwett and Rigotti 2019); irritation (Baethge et al. 2019), job resources (Afsharian et al. 2018)] and outcomes [e.g., organizational citizenship behaviors (Smith et al. 2020); turnover intentions (Steffens et al. 2018)] denoted within the stressor taxonomy framework (Halbesleben 2010). Furthermore, leveraging the larger heterogeneous nature of Sample 3, we also examined issues of measurement invariance [i.e., based on gender, age, employment status, work schedule, and tenure].

Sample 1: Methods

Participants & Procedures Data were collected online from 27 SMEs. SMEs included faculty ($n = 11$), doctoral candidates ($n = 14$), and applied practitioners ($n = 2$) in industrial/organizational psychology. SMEs had, on average, 6.04 years of experience in the field ($SD = 4.54$), 66.7% of SMEs indicated they were either *moderately* or *extremely* familiar with the construct; 51.8% had either published or presented scholarly work related to engagement and/or consulted with organizations on issues related to engagement.

SMEs were given the UWES-9 (items presented in random order) and construct definitions for both engagement and its three components. As recommended (Hinkin and Tracey 1999), SMEs were asked to assign each item to the characteristic it best represented; SMEs could not assign a given item to multiple characteristics. Then, on separate pages, SMEs reviewed the construct definitions again, along with the three items representing each engagement characteristic, and were asked to rank the items in terms of which best represented each specific definition (i.e., SMEs were asked to pick the best overall item for that characteristic, and so forth).

Sample 1: Results & Discussion

Results are presented in Table 1. Specific to vigor, item 2 was classified correctly 100% of the time, whereas item 1 was ranked the highest in terms of the degree to which it represented that dimension. For dedication, item 1 was ranked highest, and classified correctly 78% of the time, whereas item three was ranked second, and classified correctly 100% of the time. For absorption, item 1 was categorized correctly 100% of the time, and also ranked the highest. Alternatively, Vigor3, Dedication2, and Absorption3, demonstrated classification inconsistency and/or low rankings. However, judgmental assessments from SMEs are just one piece of information when making

Table 1 Study 1 sample 1 subject matter expert evaluations

UWES-9	Correctly classified	Rank
Vigor 1: At my work, I feel bursting with energy	93%	1.59
Vigor 2: At my job, I feel strong and vigorous	100%	1.96
Vigor 3: When I get up in the morning, I feel like going to work	48%	2.44
Dedication 1: I am enthusiastic about my job	78%	1.78
Dedication 2: My job inspires me	93%	2.25
Dedication 3: I am proud of the work that I do	100%	1.96
Absorption 1: I am immersed in my work	100%	1.33
Absorption 2: I get carried away when I am working	96%	1.96
Absorption 3: I feel happy when I work intensely	70%	2.69

decisions about items to potentially retain or remove from a measure (Stanton et al. 2002). As such, we retained all items as part of our examination of internal and external qualities in Samples 2 and 3 in order to develop a more holistic understanding of the items and their strengths and weaknesses.

Sample 2: Methods

With Sample 2 we evaluated internal and external qualities of the items (Stanton et al. 2002). Internal qualities were examined in terms of factor loadings via confirmatory factor analysis (CFA) as well as issues related to residual error (Smith et al. 2000). Specific to external characteristics we leveraged two taxonomies to select constructs to examine the nomological network around the items. Based on Sonnentag and Frese's (2003) stressor taxonomy, we selected constructs to index four classes of stressors: *task related job stressors*, *role stressors*, *social stressors*, and *schedule related stressors*. We also examined (physical, affective, behavioral) outcomes of engagement indexed based on Weiner et al. (2012).

Participants & Procedures Participants were recruited from [Amazon.com](https://www.amazon.com)'s Mechanical Turk (Mturk) as part of a larger effort (Matthews and Ritter 2016). Only U.S. participants who had previously completed at least 500 tasks with a 96% approval rating (as tracked and managed by Mturk) were permitted to participate. Respondents were required to work at least 24 hours a week, with no more than 50 % of their work being done at home, and were paid \$2.50 for participating. Consistent with both emerging recommendations regarding the use of online panels (Huang et al. 2012) and past research (e.g., McGonagle et al. 2016), respondents who failed to correctly complete at least five of six effortful responding questions (e.g., “leave this question blank”) were excluded. We allowed respondents to miss one attention check item given respondents may mistakenly miss one item but still, generally, be attentive (Huang et al. 2012; McGonagle et al. 2016).

A total of 299 people participated; 24 were excluded for not meeting inclusion criteria (e.g., a screener question asked the number of hours per week participants worked, wherein we excluded respondents who reported they worked less than 24 hours per week). Respondents were blind to our analysis inclusion criteria. To facilitate interpretation, we employed listwise deletion at the item level for the UWES-9 items, resulting in the exclusion of nine additional respondents. The analysis sample ($N = 264$) was 53.8% female, primarily Caucasian (76.9%) with an average age of 34.44 years ($SD = 10.51$) and organizational tenure of 4.98 years ($SD = 4.77$). On average, respondents worked 40.69 h a week ($SD = 6.22$) and 70.2% worked a day shift.

Measures Constructs selected to index the classes of stressors and outcomes are reported in Table 2. Participants were asked to consider the past month while responding to all items.

Table 2 Study 1, samples 2 and 3, stressor and outcome indices

	Sample	Indexed construct	# items	α	Response scale	Citation
Stressors						
Task-related job stressors	2	Job autonomy	3	.87	5 pt. agreement	Morgeson and Humphrey (2006)
	3	Job resources	6	.84	5 pt. agreement	Rousseau and Aubé (2010)
Role stressors	2	Role overload	5	.89	5 pt. agreement	Matthews et al. (2010)
	3	Work frustration	3	.77	5 pt. agreement	Peters et al. (1980)
Social stressors	2	Abusive supervision	5	.93	5 pt. agreement	Mitchell and Ambrose (2007)
	3	Diversity climate	9	.96	5 pt. meets expectations	McKay et al. (2007)
Schedule related stressors	2	Shift	1	–	1-day shift	2-non-standard
	–	Average work hours	1	–	–	–
3						
Outcomes						
Physical	2	Energy	3	.65	7 pt. energy index	Britt et al. (2013)
	3	Ego depletion	5	.87	5 pt. frequency	Lin and Johnson (2015)
Affective	2	Affective commitment	3	.92	5 pt. agreement	Griffen et al. (2007)
	3	Workplace anger	6	.89	5 pt. frequency	Watson and Clark (1994)
Behavioral	2	Turnover intentions	3	.93	5 pt. agreement	Kelloway et al. (1999)
	3	Organizational citizenship behaviors	14	.76	5 pt. frequency	Williams. and Anderson (1991)

Sample 2: Results & Discussion

Item-level means and standard deviations for the UWES-9 are reported in Table 3. To understand the internal qualities of the items, we first examined the factor loadings of the items from the UWES-9 based on a three-factor CFA (AMOS-21; Arbuckle 2012). Items were loaded on their respective latent factor, and the three (latent) characteristics were set free to correlate. The model fit the data well [$\chi^2(24) = 48.97$, $p < .001$, CFI = .99, RMSEA = .06, SRMR = .03] with the three factors correlating strongly ($r = .82$ to $.93$, $p < .001$). Table 3 includes standardized loading and modification indices, which were examined for issues of residual error.¹ The strongest loading items for each characteristic were Vigor2, Dedication1, and Absorption1. Vigor3 demonstrated the most systematic residual error with other items in the model. We also examined a single factor model which fit the data [$\chi^2(27) = 108.91$, $p < .001$, CFI = .95, RMSEA = .10, SRMR = .04] but demonstrated worse fit than the 3-factor model [$\Delta\chi^2(3) = 59.94$, $p < .001$].

Based on Matthews and Ritter (2016), forward statistical regression was used to examine external qualities of the different engagement items to understand which items demonstrated unique shared variance with the selected indices. By using forward statistical regression, redundant items (items that explained limited incremental variance) were identified (Stanton et al. 2002) and excluded from the statistical model. Results for the four stressor index constructs are reported in Table 4, and results for the three outcome index constructs are reported in Table 5 (coefficients are standardized beta weights).

These results provided additional item selection guidance. Specifically, Dedication1 not only demonstrated the highest factor loading (i.e., internal criteria), but was also uniquely related to six of the seven index constructs. Specific to vigor and absorption, the data are less intuitive, especially if Sample 1 SME evaluations are taken into consideration. While Vigor2 demonstrated the highest factor loading, it was systematically related to only one of the index constructs. On the other hand, Vigor3 was uniquely related to four of the seven index criteria. However, based on SME evaluations, this item had poor conceptual representation of the construct. Further, Vigor3 demonstrated systematic residual error variance with five of the other engagement items. Finally, while Absorption1 demonstrated the strongest factor loading and had strong SME evaluations, it explained no unique variance in the index constructs. If nothing else, as it relates to the larger measure adaptation literature (e.g., Heggstad et al. 2019), it is clear that no one piece of evidence (i.e., judgmental, internal, or external) is sufficient when making decisions about items to retain and exclude – a balance act exists between the varying pieces of evidence.

¹ The modification indices, and by extension the residual error they represent, are indicative of model misfit because the noted items share residual variance not accounted for in the measurement model. These suggested modifications (i.e., correlating of error terms for given items) were not estimated as part of the model testing process and are only reported for diagnostic purposes (Stanton et al. 2002).

Table 3 UWES-9 item level psychometric information to study 1, sample 2 and 3

Item	Study 1 sample 2 (N = 264)				Study 1 sample 3 (N = 1098)			
	Mean	SD	Loadings	Residual error observed	Mean	SD	Loadings	Residual error observed
Vigor1	3.02	1.13	.84		3.11	0.98	.78	Dedication3, Vigor2
Vigor2	3.34	1.07	.85	Vigor3	3.49	0.96	.77	Vigor1, Vigor3
Vigor3	3.04	1.17	.80	Vigor2, Absorption1,3, Dedication2,3	3.14	1.09	.76	Vigor2
Dedication1	3.33	1.18	.89		3.67	1.01	.88	
Dedication2	3.12	1.19	.86	Vigor3	3.49	1.06	.83	
Dedication3	3.74	1.03	.80	Vigor3	4.12	0.86	.68	Absorption1, Vigor1
Absorption1	3.60	1.04	.84	Vigor3	3.60	0.96	.73	Absorption2, Absorption3, Dedication3
Absorption2	3.20	1.04	.75		3.25	1.04	.52	Absorption1
Absorption3	3.66	1.03	.68	Vigor3	3.73	0.93	.66	Absorption1

Sample 3: Methods

Following these preliminary results, Sample 3, a larger heterogeneous sample, was used to evaluate additional internal and external evidence to further inform item

Table 4 Results of forward statistical regressions based on study 1 stressor indices, samples 2 and 3

Predictors	Study 1 sample 2 (N = 264)				Study 1 – sample 3 (N = 1098)			
	Job autonomy	Role overload	Abusive supervision	Shift	Job resources	Work frustration	Diversity climate	Hours
Vigor1						-.13**	.09*	
Vigor2					.12**	-.21**	.08*	-.11**
Vigor3		-.34**				-.18**		.08*
Dedication1	.21**	-.16*	-.30**	.22**	.29**	-.19**	.17**	
Dedication2				-.30**			.12**	
Dedication3	.29**				.09*	.09**		
Absorption1					-.12**	.16**		.19**
Absorption2						.12**		
Absorption3	.20**				.08*			
<i>F</i>	49.14**	34.55**	26.26**	4.57**	46.51**	41.49**	33.98**	15.03**
<i>R</i> ²	.36	.21	.09	.03	.18	.21	.17	.04
DF	3, 260	2, 260	1, 262	2, 258	5, 1089	7, 1083	4, 1087	3, 1067

**p* < .05

***p* < .01

Table 5 Results of forward statistical regressions based on study 1 outcome indices, samples 2 and 3

Predictors	Study 1 – sample 2 (N = 264)			Study 1 – sample 3 (N = 1098)		
	Energy	Affective commitment	Turnover intentions	Depletion	Work related anger	Organizational citizenship behaviors
Vigor1	.22**			-.20**		
Vigor2				-.13**	-.15**	
Vigor3	.17*	.15*	-.16*	-.21**	-.22**	
Dedication1	.17*		-.23**		-.16**	.18**
Dedication2		.44**	-.28**		-.10*	.11**
Dedication3		.20**				
Absorption1						.10**
Absorption2				.19**	.15**	
Absorption3						
<i>F</i>	30.06**	85.60**	51.71**	61.34**	71.20**	41.52**
<i>R</i> ²	.26	.50	.37	.18	.25	.13
<i>DF</i>	3, 263	3, 263	3, 260	4, 1093	5, 1087	4, 1089

* $p < .05$ ** $p < .01$

selection. Factor loadings and residual errors were again used to examine item internal qualities. For external characteristics, we again leveraged the taxonomies by Sonnentag and Frese (2003) and Weiner et al. (2012), but different index constructs were selected to expand the nomological network under consideration and provide further evidence of concurrent validity.

Participants & Procedures Participants were recruited using a peer-nomination web-based survey: 133 trained undergraduate students from eight geographically dispersed universities were trained on ethics relating to participant recruitment and inclusion criteria. Students distributed a standardized email invitation to working adults they personally knew and received nominal course credit. Email recipients were asked to complete the 15-min online survey; participation was voluntary. Participants were entered into a drawing for a nominal online retailer gift certificate. A total of 1132 respondents completed the survey. To facilitate analyses, and consistent with Sample 2, we excluded participants with missing engagement data ($n = 34$). The final sample ($N = 1098$) was 62.8% female and primarily Caucasian (63.7%) with an average age of 37.03 years ($SD = 13.70$) and organizational tenure of 6.51 years ($SD = 7.84$). On average, respondents worked 39.98 h/week ($SD = 12.94$).

Measures Constructs selected to index stressors and outcomes are reported in Table 2. Participants were asked to consider the past month while responding to all items. We also collected demographic data on gender, age, employment status, work schedule, and tenure.

Sample 3: Results & Discussion

Item-level statistics for the UWES-9 are reported in Table 3. We again examined item factors loadings for the UWES-9 based on a three-factor CFA (AMOS 21; 2012); the model fit the data well [$\chi^2(24) = 132.94$, $p < .001$, CFI = .98, RMSEA = .06, SRMR = .03] with the three factors correlating strongly ($r = .86$ to $.90$, $p < .001$). Standardized factor loadings are reported in Table 3, as are modification indices. We again examined a single factor model. This model fit the data [$\chi^2(27) = 257.66$, $p < .001$, CFI = .95, RMSEA = .09, SRMR = .04], however demonstrated worse fit than the 3-factor model [$\Delta\chi^2(3) = 124.72$, $p < .001$].

We again used forward statistical regression; stressor results are reported in Table 4, and outcomes results in Table 5. Dedication1 again demonstrated the highest factor loading and was uniquely related to five of the index constructs. While Vigor1 demonstrated a slightly higher factor loading than Vigor2, Vigor2 still loaded highly and related systematically to five of the index constructs. Absorption1 again demonstrated the strongest factor loading and it related uniquely to four of the index constructs.

Short-Form Item Selection Collectively, based on the three independent samples, wherein we leveraged systematically different methodologies and respondents, we identified three items for inclusion in the short-form engagement measure. While some subjectivity is involved here, our decisions were based on the observed (i.e., objective) data. To be as transparent as possible (and address potential concerns regarding subjective item selection; Heggstad et al. 2019), Table 6 includes the selected items and a summary of retention rationale for each. The abridged measure demonstrated adequate internal consistency (Sample 2, $\alpha = .85$; Sample 3, $\alpha = .79$) and correlated

Table 6 Justification for final item selection for the three-item measure of engagement

Characteristic	Item	Justification
Vigor	At my job, I feel strong and vigorous (Vigor 2)	Consistently classified by SMEs, with a relatively high ranking Highest factor loading in Sample 2 and second highest in Sample 3 Most systematically related to index constructs based on Sample 3
Dedication	I am enthusiastic about my job (Dedication 1) ^a	Highest SME ranking as well as consistent classification Highest factor loading in both Sample 2 and 3 Most systematically related to index constructs in both Sample 2 and 3
Absorption	I am immersed in my work (Absorption 1) ^a	Consistently classified by SMEs, with the highest ranking Highest factor loading in both Sample 2 and 3 Most systematically related to index constructs based on Sample 3

^a Consistent with Schaufeli et al. (2017) and Reina-Tamayo et al. (2017) short-form measures

strongly with the UWES-9 overall (Sample 2, $r = .95$; Sample 3, $r = .94$, $p < .001$). Moreover, each item correlated strongly with the longer subscale of its respective dimension (Sample 2, .88–.92; Sample 3, .81–.90). These results suggest that not only is the abridged measure internally consistent, but it also overlaps almost entirely with the UWES-9: Even with strategically removing two-thirds of the items, we still effectively capture the overall intended construct space (Smith et al. 2000). Of note, readers intending to use this short-form measure in practice should request permission from the authors of the original UWES (Schaufeli et al. 2002).

In terms of convergent validity, Table 7 reports bivariate correlations between the short- and long-form engagement measures and the predictor and outcome variables used in the item selection process for Sample 3. The bivariate correlations are consistently smaller for the short-form measure, however, the pattern of relationships between antecedents and outcomes is generally the same for both versions. In fact, average bivariate correlations for the short-form measure is .017 smaller than the long-form resulting in an average difference in variance explained of .0003; per Smith et al. (2000) we would argue this is an acceptable reduction in validity relative to the number of items removed from the long-form measure.

Multiple-Groups Invariance Testing Based on Sample 3 data, we conducted a series of multiple-groups CFA invariance tests based on gender (men = 1, women = 2), age (1 = 18 to 27, 2 = 28 to 45, 3 = 46+ years of age), employment status [1 = Part-time work (i.e., less than 35 h/week), 2 = full-time work (35+ hours/week)], work schedule (1 = day shift, 2 = alternative shift schedule), and tenure (1 = less than 1 year, 2 = 1–1.99 years, 3 = 2–4.9 year, 4 = 5–9.9 years, 5 = 10+ years). For age and tenure, groups were recoded to produce approximate equal group sizes to facilitate analyses. Hours worked was coded based on the Bureau of Labor Statistics 35-h definition for full-time employment.

Table 7 Study 1 Sample 3 convergent validity evidence: comparing the short-form to the long-form measure

Variables	Mean	SD	1	2	3	4	5	6	7	8	9
1. Engagement (short-form)	3.46	0.82	(.79)								
2. Engagement (long-form)	3.51	0.74	.94**	(.90)							
3. Job resources	3.87	0.71	.33**	.36**	(.84)						
4. Work frustration	2.89	0.94	-.31**	-.33**	-.35**	(.77)					
5. Diversity climate	3.29	0.89	.36**	.37**	.43**	-.27**	(.96)				
6. Hours	39.98	12.94	.12**	.14**	-.08**	.12**	.01	–			
7. Depletion	2.59	0.89	-.30**	-.32**	-.20**	.42**	-.15**	-.02	(.87)		
8. Work related anger	1.86	0.83	-.41**	-.42**	-.35**	.50**	-.30**	.02	.39**	(.89)	
9. Organizational citizenship behaviors	3.84	0.55	.32**	.33**	.19**	.00	.23**	.14**	-.07*	-.07*	(.76)

* $p < .05$

** $p < .01$

A separate multiple-groups CFA was conducted for each grouping variable based on recommendations by Vandenberg and Lance (2000).² We first estimated an unconstrained model, subsequently testing for metric invariance wherein we constrained item factor loadings to be equal across groups. Next, we tested for scalar invariance by constraining item intercepts to be equal across groups. Finally, we tested for invariant uniquenesses by constraining item variance to be equal across groups. As reported in Table 8, each model demonstrated acceptable fit suggesting that the short-form measure demonstrates strict factorial invariance based on gender, age, employment status, work schedule, and tenure. Complete results, including full correlation tables by grouping variable, available upon request.

Study 2

Study 2 was conducted to make two additional key contributions. First, across all three samples in Study 1, data were collected in conjunction with the excluded items (i.e., all items were administered simultaneously). As noted by Smith et al. (2000) when establishing validity evidence for a short-form measure “key empirical evidence should not be based [strictly] on a sample in which the full, long-form was administered” (p. 107). Independent tests of the validity of a short-form measure are needed to ensure the measure continues to function as expected. This is an issue Schaufeli et al. (2017, p. 13) highlight such that their measure was “not independently used from the UWES-9, so that its true reliability and validity is not yet fully understood” (p. 13). Our short-form development process overcomes this limitation.

Furthermore, an underlying argument we have put forth is that a valid short-form measure would be particularly advantageous in advanced research designs, including true longitudinal research (i.e., where data is collected on all constructs at a minimum of at least three time points; Ployhart and Vandenberg 2010). To this end, we administered the short-form measure in a longitudinal data collection (six waves of data with 2-week lags) to examine whether the measure continued to demonstrate acceptable psychometric characteristics (e.g., internal consistency) as well as additionally assessing whether the measure demonstrates temporal invariance and predictive validity.

Specifically, we focused our analyses on understanding the dynamic relationship between engagement and well-being. We frame our hypotheses within the context of gain spirals as defined by conservation of resources theory (Hobfoll 1989). Furthermore though, the majority of the existing research examining related issues has been based on cross-sectional (e.g., Steele et al. 2012) or cross-lagged panel models (Hakanen and Schaufeli 2012), all of which take a between-person analysis approach. While between-person analyses are helpful, particularly cross-lagged panel models (Selig and Little 2012), we seek to contribute to the literature by examining if an engagement/well-being gain spiral exists based on more stringent within-person analyses.

² To prevent a just-identified model (i.e., three item CFA models) we included income as an indicator variable. At an omnibus level, income correlated .08 with the short-form engagement measure, $p < .01$.

Table 8 Multiple-groups CFA invariance testing

Grouping variable	χ^2	DF	CFI	RMSEA	SRMR	$\Delta\chi^2$	ΔDF
Gender							
Unconstrained	11.29*	4	.993	.041	.026	–	–
Metric invariance	15.10*	6	.991	.037	.029	3.87	2
Scalar invariance	19.13*	9	.990	.032	.029	4.03	3
Invariant uniquenesses	21.91*	12	.990	.028	.032	2.78	3
Age							
Unconstrained	6.25	6	1.000	.006	.023	–	–
Metric invariance	16.55	10	.993	.025	.024	10.30*	4
Scalar invariance	57.17**	16	.957	.049	.024	40.62**	6
Invariant uniquenesses	69.82**	22	.950	.045	.027	12.65*	6
Hours							
Unconstrained	5.86	4	.998	.021	.006	–	–
Metric invariance	9.05	6	.997	.022	.019	3.19	2
Scalar invariance	32.70**	9	.975	.050	.024	23.65**	3
Invariant uniquenesses	36.30**	12	.945	.061	.056	3.60	3
Schedule							
Unconstrained	7.81	4	.996	.030	.022	–	–
Metric invariance	8.31	6	.998	.019	.022	0.50	2
Scalar invariance	15.77	9	.993	.027	.022	7.46	3
Invariant uniquenesses	16.47	12	.995	.019	.022	0.70	3
Tenure							
Unconstrained	13.19	10	.997	.017	.036	–	–
Metric invariance	22.28	18	.995	.015	.040	9.09	8
Scalar invariance	54.77**	30	.974	.028	.040	32.49*	12
Invariant uniquenesses	71.75**	42	.968	.026	.047	16.98*	12

* $p < .05$ ** $p < .01$

By way of background, it is generally supported that more engaged workers report higher indices of well-being, including life satisfaction. A common argument is that engagement results in improved well-being (over time) through mechanisms such as gain spirals in personal resources, wherein conservation of resources is often evoked as the driving theoretical framework (Hakanen and Schaufeli 2012; Hobfoll 2011). Specific to life satisfaction, a quintessential index of well-being (Diener and Ryan 2009), as noted, much of the research supporting this claim has been cross-sectional (e.g., Körner et al. 2012; Matthews et al. 2014; Steele et al. 2012). Existing over-time studies also support this conclusion (e.g., Hakanen and Schaufeli 2012; Reis et al. 2016). However, as noted, we take a more stringent approach to our analyses, applying a lagged fixed-effects structural equation model to examine whether, and the degree to which, changes in engagement predict future (lagged) changes in satisfaction, for a given person.

- *Hypothesis 1: Within-person changes in engagement predict lagged within-person changes in life satisfaction.*

Interestingly though, consistent with the larger stressor-strain literature (Ford et al. 2014), examination of potential reverse causality (i.e., that life satisfaction predicts engagement) has been limited. Nevertheless, an argument for reverse causation can be grounded in the same gain spirals argument put forth in extant research (e.g., Hakanen and Schaufeli 2012). That is, as a proxy for well-being, higher life satisfaction suggests that an individual has a larger pool of personal resources to draw from, across *multiple* life domains (Diener and Diener 1996). In turn, this personal resource (i.e., great life satisfaction) can in turn be invested to gain further resources, for example, at work (Hobfoll 2011). Put into the context of this study, and consistent with Principle 2 as well as Corollary 1 of conservation of resource theory (Hobfoll 1989, 2011), having more generalized resources, like life satisfaction, should facilitate accumulation of more domain-specific resources, like work engagement.

- *Hypothesis 2: Within-person changes in life satisfaction predict lagged within-person changes in engagement.*

Method

Participants & Procedures

Participants were recruited from [Amazon.com](https://www.amazon.com)'s MTurk. Only U.S. participants who had previously completed at least 100 tasks with a 98% approval rating were permitted to participate. While Study 1 Sample 2 respondents were excluded from participating we followed a similar screening method; respondents were required to work at least 24 h a week, be organizationally employed, with no more than 40% of their work being done at home. Participants were asked to complete the same survey six times, with a 2-week lag between assessments. Reminder emails were sent approximately 4 days after the initial invitation, at each wave of data collection. Five validation questions (e.g., "Please leave this item blank") were included to ensure effortful responding (again, participants were allowed to miss one attention check item; Huang et al. 2012; McGonagle et al. 2016). Participants were paid \$1.20 for the first survey, and \$1.00 for each of the remaining surveys.

A total of 1506 participants were screened at Time 1. Of these, 667 did not meet our inclusion criteria and were excluded from participating; another 31 failed more than one of the attention check items. The remaining 808 were invited back to complete the remaining five surveys. Response rates for the remaining five surveys ranged between 68.2% and 81.8%. To be retained for analyses respondents had to complete at least two of the six surveys, remain employed over the course of the study with no major job changes, and demonstrate effortful responding on the remaining surveys. This resulted in an analysis sample of 627 which was 49.9% female, primarily Caucasian (75.9%) with an average age of 36.98 years ($SD = 10.41$) and position tenure of 5.24 years ($SD = 4.89$). On average, respondents worked 41.69 hours a week ($SD = 6.77$) and 74.0% worked a day shift. Of note, within the analysis sample, on average, participants

completed 5.21 ($SD = 1.01$) of the six surveys. Sample sizes, by survey wave, were as follows: Time 2 = 562; Time 3 = 555; Time 4 = 523; Time 5 = 496; Time 6 = 504.

Measures

Engagement was assessed with the short-form measure developed in Study 1 (see Table 6). *Life satisfaction* was measured with a 5-item measure (Diener 1984). A sample item is, “In most ways my life is close to my ideal.” For both measures, participants were asked to consider the past 2 weeks when responding; responses were on a 5-point agreement-based Likert scale.

Results & Discussion

Table 9 reports descriptive statistics. Independent of the excluded items from the larger original-form (Smith et al. 2000) the short-form measure demonstrated good internal consistency at all six time points (as did the life satisfaction measure).

Table 9 Study 2 descriptive statistics

Variables	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. T1 engagement	3.52	0.89	(.85)											
2. T1 life satisfaction	3.29	0.96	.43	(.90)										
3. T2 engagement	3.49	0.89	.74	.40	(.84)									
4. T2 life satisfaction	3.35	0.96	.41	.89	.42	(.92)								
5. T3 engagement	3.48	0.91	.74	.40	.77	.42	(.85)							
6. T3 life satisfaction	3.35	0.98	.45	.85	.46	.91	.50	(.93)						
7. T4 engagement	3.49	0.92	.70	.40	.76	.43	.76	.49	(.86)					
8. T4 life satisfaction	3.35	0.98	.41	.87	.39	.92	.44	.93	.47	(.92)				
9. T5 engagement	3.46	0.91	.70	.39	.74	.42	.79	.48	.76	.42	(.87)			
10. T5 life satisfaction	3.34	0.98	.41	.86	.41	.92	.42	.91	.45	.92	.47	(.92)		
11. T6 engagement	3.45	0.97	.68	.42	.76	.44	.74	.50	.76	.43	.76	.47	(.90)	
12. T6 life satisfaction	3.34	0.99	.43	.86	.43	.91	.45	.90	.47	.91	.48	.93	.50	(.93)

Note: T Time. All correlations significant at $p < .01$

Temporal Invariance Testing

Next, we examined both measures for temporal invariance (Ployhart and Vandenberg 2010). We conducted baseline confirmatory factor analyses (CFA) for each construct separately. Within each CFA, items were used as indicators at each measurement assessment (Time 1 – Time 6). The six latent factors were loaded onto a second order latent factor. In turn we sequentially tested for configural (i.e., loadings on the second order factor were constrained to be equal), metric (i.e., item loads across first order latent factors were constrained to be equal), and scalar invariance (i.e., item intercepts were constrained to be equal), as well as for invariant uniquenesses (i.e., item uniquenesses were constrained to be equal; Vandenberg and Lance 2000). While full results are available upon request, the final constrained model for the short-form engagement measure fit the data [$\chi^2(128) = 207.28, p < .001, CFI = .99, RMSEA = .03$]; the model for life satisfaction did as well [$\chi^2(398) = 703.52, p < .001, CFI = .99, RMSEA = .04$]. Thus, both constructs demonstrated strict factorial invariance.

Examination of Temporal Order

To examine the potential for dynamic interplay between engagement and life satisfaction, as predicted by conservation or resources theory and represented by Hypotheses 1 and 2, we applied a latent variable fixed-effects estimator approach following recommendations by Allison (2000, 2005) with maximum likelihood estimation. A fixed effects method allows us to control for all unchanging (time invariant) variables (e.g., personality, job design) and removes all between-individual variation. In effect, not only does this increase our confidence that the lagged (causal) relationship between engagement and life satisfaction is not the result of some extraneous variable, it also allows for examination of the effect in terms of within-individual changes (which is facilitated by our establishment of strict factorial invariance for both measures; Ployhart and Vandenberg 2010). Further, our large sample size (Ford et al. 2014) and overall number of assessments (allowing for repeated estimation of lagged effects) helps ensure the observed relationships are not due to chance. Collectively, while not a randomized experiment, our approach affords a strong case for understanding causal order between the constructs (Shadish et al. 2002).

Given that both measures demonstrated strict temporal invariance, to reduce model complexity, constructs were modeled as directly observed (i.e., scale scores were used). While a full discussion of the set-up and estimation process for latent variable fixed-effects estimating within SEM is beyond the scope of this paper, interested readers are encouraged to see Ousey et al. (2011) for an applied example with an expanded discussion of Allison's (2000, 2005) recommendations. However, the process can be effectively depicted, and to this end, unstandardized results for the latent variable fixed-effects model (and associated constraints) are reported in Fig. 1. As depicted, autoregressive and lagged effects (i.e., coefficients that share a superscript) were constrained to be equal across assessments to further reduce model complexity; to facilitate interpretation, correlations (depicted as double headed arrows) are reported. The model fit the data well [$\chi^2(49) = 72.68, p = .02, CFI = .997, RMSEA = .028, SRMR = .025$].

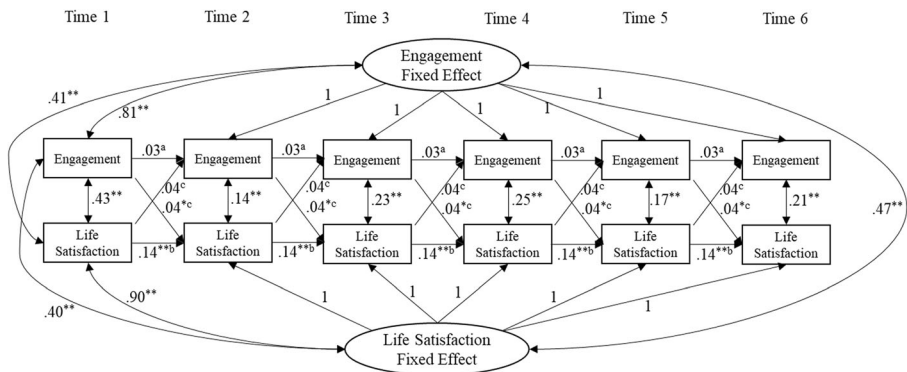


Fig. 1 Unstandardized results for Study 2 latent variable fixed-effect model examining causal order between engagement and life satisfaction. Coefficients sharing a superscript were constrained to be equal across lags. Correlations are reported for all double-headed arrows

In addition to testing our hypotheses, our analyses suggest several noteworthy issues. First, the non-significant autoregressive effect for engagement ($B = .03$, $p > .05$) suggests that once time invariant predictors are controlled for (e.g., personality, job design) there is no lagged state dependency for engagement. Put another way, the average autoregressive correlation reported in Table 9 is .74. Once time invariant predictors are accounted for, there is no evidence of an over-time relationship in engagement with itself; for a given individual, a change in engagement (increase/decrease) does not result in a subsequent change in engagement 2 weeks later for that individual. However, there is evidence supporting state dependency for life satisfaction. A one unit increase in life satisfaction resulted in a .14 increase in life satisfaction 2 weeks later.

Specific to our hypotheses, we see a small, albeit significant, lagged effect between engagement and life satisfaction ($B = .04$, $p < .05$); Hypothesis 1 was supported. However, we do not see a lagged effect of life satisfaction on engagement ($B = .04$, $p > .05$); Hypothesis 2 was not supported. Collectively then, and in the context of conservation resources theory as well as the larger well-being literature (e.g., Headey et al. 1991), our results support what might be termed a *bottom-up gain spiral* in terms of engagement and life satisfaction. That is, increases in engagement drove lagged increases in life satisfaction; gaining resources at work (in the form of engagement) resulted in respondents (based on our within-person analysis) reporting experiencing more generalized resources (in the form of increased life satisfaction). However, we did not observe evidence for a *top-down gain spiral*. Specifically, increases in generalized resources (i.e., life satisfaction) did not drive changes in, or rather the accumulation of resources, in the work domain (in terms of changes in engagement). We would suggest then that, given the methodological and analytical approach used here, both of which were facilitated by the application of our short-form measure, we have a clearer understanding of the lagged effects of engagement on life satisfaction.

General Discussion

The studies reported herein are informed by understanding the value in formulating a rigorously validated measure of engagement to facilitate more coalesced research and practice. We present extensive evidence for the efficacy of our short-form measure across two studies and multiple data sources, using a multifaceted approach to validation. Collectively, we demonstrate that the short-form engagement measure is psychometrically sound in and of itself, as well as evidencing that it functions similarly to its parent item set. Subsequently, so as to further inform an appropriate understanding of this construct and ensure its representative measurement, and address the existing void in the literature, we established that the short-form measure is both invariant across several demographic characteristics, but is also temporally invariant, a critical precondition for scholars seeking to understand “changes” in engagement over time. We elaborate on these issues next.

Theoretical and Practical Implications

Our research has important implications for engagement operationalization and measurement, both in and of itself as well as how it relates to other important constructs both within and over time. Importantly, we found expected relations with relevant constructs, supporting concurrent validity (Study 1, Sample 3), as well as a more in-depth evaluation of engagement alongside life satisfaction (Study 2), filling a gap in the literature that has heretofore failed to examine the extent to which work engagement may impact employees’ broad well-being indices over time (or vice-versa). This has implications for how we understand the engagement construct conceptually, as well as for the optimal content design of future engagement research.

Our findings also have tangible practical implications insofar as optimizing utilitarian and psychometrically sound engagement measurement. In Study 1, we provided strong empirical and theoretical justification for a valid three-item measure of engagement that, while derived from the most popular research-based engagement instrument, is also parsimonious enough for implementation in practice-based efforts. Such a contribution takes an important step forward in closing the research-practice gap (e.g., Saks and Gruman 2014; Shuck et al. 2017). In presenting such an instrument that can be used across research and practice domains alike, as well as across a variety of research designs (cross-sectional, longitudinal, frequent administrations, etc.), we equip scholars and practitioners with a crucial tool. Such a tool is necessary to align and thereby lay the groundwork for more research-informed engagement practice within academia and practice. In particular, the parsimony of this measure, combined with its theoretical soundness and empirical support, makes it more ubiquitously attractive to a wide variety of researchers and practitioners alike seeking to assess engagement. As such, this measure facilitates consistency and generalizability across studies, as well as benchmarking across organizations and industries, thereby allowing for more accurate comparative assessments than are currently provided by different and inconsistent measures of varying lengths and item texts – thus offering all those seeking to assess engagement a ‘common language,’ as we noted earlier, with which to do so.

Yet another practical implication of this research is the examination of the instrument’s measurement invariance across a number of salient demographic characteristics

relevant in the organizational sciences. With both researchers and practitioners increasingly seeking to make comparisons across grouping variables (e.g., men vs women, 20-somethings vs baby boomers), it is critical to establish that any measure used in that way does not function differentially across the groups in question. As such, the present research's determination that our short-form measure functions invariantly across the critical demographic characteristics – within a large heterogeneous sample – allows researchers and practitioners alike to make comparisons across these groups with an improved level of confidence.

The present series of studies is also worthwhile to furthering engagement research on the whole, as well as optimizing its contextualization within the occupational health literature (Bakker et al. 2008). That is, whereas a majority of engagement research is cross-sectional or otherwise fails to account for the invariant nature of the construct over time (e.g., Kim and Beehr 2020), much of that research simultaneously suggests that improving engagement is critical because, among other things, engagement begets engagement in a sort of positive gain spiral such as that put forth in conservation of resources theory. However, preliminary evidence from Study 2 suggests that these are more complex issues than initially recognized. First, we do see evidence for a bottom-up gain spiral; changes in engagement drove (lagged) changes in life satisfaction. That is, gaining resources at work (i.e., becoming more engaged) resulted in perceived changes in generalized resources (i.e., life satisfaction). However, gaining more generalized resources (i.e., life satisfaction) did not drive domain specific resource accumulation in the form of (lagged) changes work engagement – there was no evidence for top-down gain spiral (Headey et al. 1991).

Yet, arguably even more interesting as related to gain spirals is how changes in engagement, at least based on our data, are not related to future changes in engagement at the within-person level. That is, once time invariant predictors are controlled for (e.g., personality, job design) there was no lagged state dependency for engagement; for a given individual, a change in engagement (increase/decrease) did not result in a subsequent change in engagement 2 weeks later for that individual. Specific to conservation of resources theory, our data does not support the argument that engagement begets more engagement over time. These results, collectively, are critical because they raise meaningful questions about the theoretical underpinnings of engagement functioning. Specifically, with conservation of resources theory serving as a popular theory in which to ground engagement research, our findings suggest there are meaningful boundary conditions around the extent to which the gain spiral central to this and other (e.g., broaden-and-build) popular positive organizational behavior theories is an optimal frame within which to view engagement.

As such, our research provides an important baseline understanding of engagement's invariant nature – an understanding critical in informing future change-focused studies of engagement, as well as practical considerations such as the value of interventions and other such attempts (e.g., job re-design, job crafting; Kuijpers et al. 2020) to promote engagement over time. In this way, longitudinal designs such as that undertaken herein are central to accurately modeling engagement's true functioning as well as the directionality of its relationship with well-being indices over time, and the short-form measure of engagement that we establish and validate herein greatly facilitates such longitudinal studies without compromising soundness of measurement.

Limitations and Future Directions

Despite the wide-reaching benefits of the short-form measure, it is important to note that the three-item measure does limit the extent to which the instrument can be partitioned into the separate engagement components (vigor, dedication, absorption). It has been argued that such an approach is not psychometrically optimal or may be variably acceptable depending upon the circumstance or construct (e.g., Fisher et al. 2016). That said, the engagement construct in particular has been the center of several assessments of factorial structure, with a number of researchers (e.g., Christian et al. 2011; Sonnentag 2003) determining that the unifactorial measure often functions better than the tripartite assessment. As such, while an important consideration for interpretation and usage, we do not believe that using the short-form measure for a unifactorial interpretation of engagement is detrimental. On the contrary, in addition to the aforementioned implications for research and practice, the measure also provides an important opportunity for future research to leverage unique designs and rigorous, longitudinal measurement that is largely impossible with the current longer engagement measures, or even with the existing abbreviated measures which lack sufficient validity evidence. In this way, this measure presents a uniquely practicable opportunity for research and practice to align in their assessment of this critical construct which has been deemed one of the most important constructs in organizational science in recent years (Saks 2017).

Conclusion

We provide extensive and multifaceted validity evidence supporting the psychometric utility of a short-form measure of engagement that aligns with the predominant operationalization of this popular construct. Not only does this short-form measure allow researchers to more confidently conduct complex designs, but its parsimonious length stands to make noteworthy strides in diminishing the ever-problematic scientist-practitioner gap regarding engagement. In this way, this instrument has the potential to begin rectifying a central concern consistently lamented by scholars – that academics and practitioners are failing to measure the same construct, and, relatedly, that an optimal measure for consistent use across research and practice does not yet exist. The measure outlined herein intends to serve that end.

References

- Afsharian, A., Zadow, A., Dollard, M. F., Dormann, C., & Ziaian, T. (2018). Should psychosocial safety climate theory be extended to include climate strength? *Journal of Occupational Health Psychology*, 23, 496–507.
- Allison, P. D. (2000). Inferring causal order from panel data. Paper prepared for presentation at the ninth international conference on panel data, 22 June. Geneva, Switzerland: SAS Institute.
- Allison, P. D. (2005). *Fixed effects regression methods for longitudinal data using SAS*. Cary: SAS Institute, Inc.
- Arbuckle, J. L. (2012). *AMOS (version 21.0) [computer software]*. Chicago: SPSS.

- Baethge, A., Deci, N., Dettmers, J., & Rigotti, T. (2019). “Some Days Won’t End Ever”: Working faster and longer as a boundary condition for challenge versus hindrance effects of time pressure. *Journal of Occupational Health Psychology, 24*, 322–332.
- Bailey, C. (2016). Employee engagement: Do practitioners care what academics have to say – and should they? *Human Resource Management Review*. <https://doi.org/10.1016/j.hrmr.2016.12.014> Advance online publication. <https://www.sciencedirect.com/science/article/abs/pii/S1053482216301061>.
- Bakker, A. B., & Leiter, M. P. (Eds.). (2010). *Work engagement: A handbook of essential theory and research*. New York: Psychology Press.
- Bakker, A. B., & Schaufeli, W. B. (2008). Positive organizational behavior: Engaged employees in flourishing organizations. *Journal of Organizational Behavior, 29*, 147–154.
- Bakker, A. B., Schaufeli, W. B., Leiter, M. P., & Taris, T. W. (2008). Work engagement: An emerging concept in occupational health psychology. *Work & Stress, 22*, 187–200.
- Bakker, A. B., Albrecht, S. L., & Leiter, M. P. (2011). Key questions regarding work engagement. *European Journal of Work and Organizational Psychology, 20*, 4–28.
- Balducci, C., Fraccaroli, F., & Schaufeli, W. B. (2010). Psychometric properties of the Italian version of the Utrecht Work Engagement Scale (UWES-9). *European Journal of Psychological Assessment, 26*, 143–149.
- Britt, T. W., McKibben, E. S., McFadden, A., & Kelley, C. (2013). *Assessing state energy as the perceived resources available to accomplish different types of tasks*. Clemson: Unpublished Manuscript, Clemson University.
- Buckingham, M., & Coffman, C. (1999). *Break all the rules*. London: Simon & Shuster.
- Byrne, Z. S., Peters, J. M., & Weston, J. W. (2016). The struggle with employee engagement: Measures and construct clarification using five samples. *Journal of Applied Psychology, 101*, 1201–1227.
- Chen, S., Shih, C., & Chi, N. (2018). A multilevel job demands-resources model of work engagement: Antecedents, consequences, and boundary conditions. *Human Performance, 31*, 282–304.
- Christian, M. S., Garza, A. S., & Slaughter, J. E. (2011). Work engagement: A quantitative review and test of its relations with task and contextual performance. *Personnel Psychology, 64*, 89–136.
- Collins, L., Fineman, D., & Tsuchida, A. (2017). People analytics: Recalculating the route. In Walsch, L. and Volini, E. (Eds.), *Rewriting the rules for digital age* (pp. 97–106). Deloitte university press.
- De Bruin, G. P., & Henn, C. M. (2013). Dimensionality of the 9-item Utrecht work engagement scale (UWES-9). *Psychological Reports, 112*(3), 788–799.
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*(3), 542.
- Diener, E., & Diener, C. (1996). Most people are happy. *Psychological Science, 7*, 181–185.
- Diener, E., & Ryan, K. (2009). Subjective well-being: A general overview. *South Africa Journal of Psychology, 39*, 391–406.
- Fisher, G. G., Matthews, R. A., & Gibbons, A. M. (2016). Developing and investigating the use of single-item measures in organizational research. *Journal of Occupational Health Psychology, 21*, 3–23.
- Ford, M. T., Matthews, R. A., Wooldridge, J. D., Mishra, V., Kakar, U., & Strahn, S. R. (2014). How do occupational stressor-strain effects vary with time? A review and meta-analysis of the relevance of time lags in longitudinal studies. *Work & Stress, 28*, 9–30.
- Griffen, M. A., Neal, A., & Parker, S. (2007). A new model of work role performance: Positive behavior in uncertain and interdependent contexts. *Academy of Management Journal, 50*, 327–347.
- Hakanen, J. J., & Schaufeli, W. B. (2012). Do burnout and work engagement predict depressive symptoms and life satisfaction? A three-wave seven-year prospective study. *Journal of Affective Disorders, 141*, 415–424.
- Halbesleben, J. R. (2010). A meta-analysis of work engagement: Relationships with burnout, demands, resources, and consequences. In A. B. Bakker & M. P. Leiter (Eds.), *Work engagement: A handbook of essential theory and research* (pp. 102–117). New York: Psychology Press.
- Halbesleben, J. R., & Wheeler, A. R. (2008). The relative roles of engagement and embeddedness in predicting job performance and intention to leave. *Work & Stress, 22*, 242–256.
- Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of Applied Psychology, 87*, 268–279.
- Headley, B., Veenhoven, R., & Wearing, A. (1991). Top-down versus bottom-up theories of subjective well-being. *Social Indicators Research, 24*(1), 81–100.
- Heggestad, E. D., Scheaf, D. J., Banks, G. C., Hausfeld, M. M., Tonidandel, S., & Williams, E. B. (2019). Scale adaptation in organizational science research: A review and best-practice recommendations. *Journal of Management, 45*, 2596–2627.

- Hinkin, T. R., & Tracey, J. (1999). An analysis of variance approach to content validation. *Organizational Research Methods*, 2, 175–186.
- Hobfoll, S. E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist*, 44(3), 513–524.
- Hobfoll, S. E. (2011). Conservation of resource caravans and engaged settings. *Journal of Occupational and Organizational Psychology*, 84, 116–122.
- Huang, J. L., Curran, P. G., Keeney, J., Potoski, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27, 99–114.
- Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, 33, 692–724.
- Kelloway, E. K., Gottlieb, B. H., & Barham, L. (1999). The source, nature, and direction of work and family conflict: A longitudinal investigation. *Journal of Occupational Health Psychology*, 4, 337–346.
- Kim, M., & Beehr, T. A. (2020). The long reach of the leader: Can empowering leadership at work result in enriched home lives? *Journal of Occupational Health Psychology*, 25, 203–213.
- Körner, A., Reitzle, M., & Silbereisen, R. K. (2012). Work-related demands and life satisfaction: The effects of engagement and disengagement among employed and long-term unemployed people. *Journal of Vocational Behavior*, 80, 187–196.
- Kronenwett, M., & Rigotti, T. (2019). When do you face a challenge? How unnecessary tasks block the challenging potential of time pressure and emotional demands. *Journal of Occupational Health Psychology*, 24, 512–526.
- Kuijpers, E., Kooij, D. T. A. M., & van Woerkom, M. (2020). Align your job with yourself: The relationship between a job crafting intervention and work engagement, and the role of workload. *Journal of Occupational Health Psychology*, 25, 1–16.
- Lapierre, L. M., Matthews, R. A., Eby, L. T., Truxillo, D. M., Johnson, R. E., & Major, D. A. (2018). Recommended practices for initiating and managing research partnerships with organizations. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 11, 543–581.
- Lin, S. H. J., & Johnson, R. E. (2015). A suggestion to improve a day keeps your depletion away: Examining promotive and prohibitive voice behaviors within a regulatory focus and ego depletion framework. *Journal of Applied Psychology*, 100, 1381–1397.
- Macey, W. H., & Schneider, B. (2008). The meaning of employee engagement. *Industrial and Organizational Psychology*, 1, 3–30.
- Matthews, R. A., & Ritter, K. J. (2016). A concise, content valid, gender invariant measure of workplace incivility. *Journal of Occupational Health Psychology*, 21, 352–365.
- Matthews, R. A., Kath, L. M., & Barnes-Farrell, J. L. (2010). Short, valid, predictive measures of work-family interference. *Journal of Occupational Health Psychology*, 15, 75–90.
- Matthews, R. A., Mills, M. J., Trout, R. C., & English, L. (2014). Family-supportive supervisor behaviors, work engagement, and subjective well-being: A contextually dependent mediated process. *Journal of Occupational Health Psychology*, 19, 168–181.
- May, D. R., Gilson, R. L., & Harter, L. M. (2004). The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work. *Journal of Occupational and Organizational Psychology*, 77, 11–37.
- McGonagle, A. K., Huang, J. L., & Walsh, B. M. (2016). Insufficient effort survey responding: An under appreciated problem in work and organisational health psychology research. *Applied Psychology: An International Review*, 65, 287–321.
- McKay, P. F., Avery, D. R., Tonidandel, S., Morris, M. A., Hernandez, M., & Hebl, M. R. (2007). Racial differences in employee retention: Are diversity climate perceptions the key? *Personnel Psychology*, 60, 35–62.
- Mills, M. J., Culbertson, S. S., & Fullagar, C. J. (2012). Conceptualizing and measuring engagement: An analysis of the Utrecht work engagement scale. *Journal of Happiness Studies*, 13, 519–545.
- Mills, M. J., Fleck, C. R., & Kozikowski, A. (2013). Positive psychology at work: A conceptual review, state-of-practice assessment, and a look ahead. *Journal of Positive Psychology*, 8, 153–164.
- Mitchell, M. S., & Ambrose, M. L. (2007). Abusive supervision and workplace deviance and the moderating effects of negative reciprocity beliefs. *Journal of Applied Psychology*, 92, 1159–1168.
- Morgeson, F. P., & Humphrey, S. E. (2006). The work design questionnaire (WDQ): Developing and validating a comprehensive measure for assessing job design and the nature of work. *Journal of Applied Psychology*, 91, 1321–1339.
- Usey, G. C., Wilcox, P., & Fisher, B. S. (2011). Something old, something new: Revisiting competing hypotheses of the victimization–offending relationship among adolescents. *Journal of Quantitative Criminology*, 27, 53–84.

- Peters, L. H., O'Connor, E. J., & Rudolf, C. J. (1980). The behavioral and affective consequences of performance relevant situational variables. *Organizational Behavior & Human Performance*, 25, 79–96.
- Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management*, 36, 94–120.
- Rath, T., & Conchie, B. (2009). *Strengths based leadership: Great leaders, teams and why people follow*. New York: Gallup Press.
- Reina-Tamayo, A. M., Bakker, A. B., & Derks, D. (2017). Episodic demands, resources, and engagement: An experience-sampling study. *Journal of Personnel Psychology*, 16, 125–136.
- Reis, D., Arndt, C., Lischetzke, T., & Hoppe, A. (2016). State work engagement and state affect: Similar yet distinct concepts. *Journal of Vocational Behavior*, 93, 1–10.
- Rich, B. L., Lepine, J. A., & Crawford, E. R. (2010). Job engagement: Antecedents and effects on job performance. *Academy of Management Journal*, 53, 617–635.
- Rousseau, V., & Aubé, C. (2010). Social support at work and affective commitment to the organization: The moderating effect of job resource adequacy and ambient conditions. *Journal of Social Psychology*, 150, 321–340.
- Saks, A. M. (2017). Translating employee engagement research into practice. *Organizational Dynamics*, 46, 76–86.
- Saks, A. M., & Gruman, J. A. (2014). What do we really know about employee engagement? *Human Resource Development Quarterly*, 25, 155–182.
- Salanova, M., Agut, S., & Peiró, J. M. (2005). Linking organizational resources and work engagement to employee performance and customer loyalty: The mediation of service climate. *Journal of Applied Psychology*, 90, 1217–1227.
- Schaufeli, W. B. (2006). The balance of give and take: Toward a social exchange model of burnout. *The International Review of Social Psychology*, 19, 87–131.
- Schaufeli, W. B. (2012). Work Engagement. What Do We Know and Where Do We Go? Work Engagement in Everyday Life, Business, and Academia. *Romanian Journal of Applied Psychology*, 14, 3–10.
- Schaufeli, W. B. (2013). What is engagement? In C. Truss et al. (Eds.), *Employee engagement in theory and practice* (pp. 15–35). New York: Routledge.
- Schaufeli, W. B., & Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *Journal of Organizational Behavior*, 25, 293–315.
- Schaufeli, W. B., & Bakker, A. B. (2010). The conceptualization and measurement of work engagement. In A. B. Bakker & M. P. Leiter (Eds.), *Work engagement: A handbook of essential theory and research* (pp. 10–24). New York: Psychology Press.
- Schaufeli, W. B., Salanova, M., Gonzalez-Roma, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A confirmative analytic approach. *Journal of Happiness Studies*, 3, 71–92.
- Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement*, 66, 701–716.
- Schaufeli, W. B., Shimazu, A., Hakanen, J., Salanova, M., & De Witte, H. (2017). An ultra-short measure for work engagement: The UWES-3 validation across five countries. *European Journal of Psychological Assessment*. <https://doi.org/10.1027/1015-5759/a000430>.
- Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 265–278). New York: Guilford Press.
- Seppälä, P., Hakanen, J., Mauno, S., Perhoniemi, R., Tolvanen, A., & Schaufeli, W. (2015). Stability and change model of job resources and work engagement: A seven-year three-wave follow-up study. *European Journal of Work and Organizational Psychology*, 24, 360–375.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton, Mifflin and Co..
- Shirom, A. (2003). Feeling vigorous at work? The construct of vigor and the study of positive affect in organizations. In Perrewe, P.L. & Ganster, D.C. (Ed.) *Emotional and Physiological Processes and Positive Intervention Strategies* (Research in Occupational Stress and Well Being, Vol. 3) pp. 135-164. Bingley: Emerald Group Publishing Limited.
- Shuck, B., Osam, K., Zigarni, D., & Nimon, K. (2017). Definitional and conceptual muddling: Identifying the positionality of employee engagement and defining the construct. *Human Resource Development Review*, 16, 263–293.
- Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, 12, 102–111.

- Smith, R. W., Kim, Y.-J., & Carter, N. T. (2020, April 9). Does it matter where you're helpful? Organizational citizenship behavior from work and home. *Journal of Occupational Health Psychology*. Advance online publication. <https://doi.org/10.1037/ocp0000>.
- Sonnentag, S. (2003). Recovery, work engagement, and proactive behavior: A new look at the interface between non-work and work. *Journal of Applied Psychology*, 88, 518–528.
- Sonnentag, S., & Frese, M. (2003). *Stress in organizations*. Konstanz: Bibliothek der Universität Konstanz.
- Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology*, 55(1), 167–194.
- Steele, J. P., Rupayana, D. D., Mills, M. J., Smith, M. R., Wefald, A. J., & Downey, R. G. (2012). Relative importance and utility of positive worker states: A review and empirical examination. *Journal of Psychology: Interdisciplinary and Applied*, 146, 617–650.
- Steffens, N. K., Yang, J., Jetten, J., Haslam, A., & Lipponen, J. (2018). The unfolding impact of leader identity entrepreneurship on burnout, work engagement, and turnover intentions. *Journal of Occupational Health Psychology*, 23, 373–387.
- Sulea, C., Virga, D., Maricutoiu, L. P., Schaufeli, W., Zaborila Dumitru, C., & Sava, F. A. (2012). Work engagement as mediator between job characteristics and positive and negative extra-role behaviors. *Career Development International*, 17(3), 188–207.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods*, 3, 4–70.
- Viljevac, A., Cooper-Thomas, H. D., & Saks, A. M. (2012). An investigation into the validity of two measures of work engagement. *The International Journal of Human Resource Management*, 23, 3692–3709.
- Watson, D., & Clark, L. (1994). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*. Ames: Unpublished manuscript, University of Iowa.
- Wefald, A. J., Mills, M. J., Smith, M. R., & Downey, R. G. (2012). A comparison of three job engagement measures: Examining their factorial and criterion-related validity. *Applied Psychology: Health and Well-Being*, 4, 67–90.
- Weiner, I. B., Schmitt, N. W., & Highhouse, S. (2012). *Handbook of psychology: Industrial and organizational psychology*. Hoboken: Wiley.
- Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*, 17, 601–617.
- Xanthopoulou, D., Bakker, A. B., Demerouti, E., & Schaufeli, W. B. (2009). Work engagement and financial returns: A diary study on the role of job and personal resources. *Journal of Occupational and Organizational Psychology*, 82, 183–200.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.