Development of an Intentional BiFactor Engagement Measure

Morgan Russell1, Casey Osorio-Duffoo2, Renata Garcia Prieto Palacios Roji1, & John Kulas1

1 Montclair State University

2 Harver

Author note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to Morgan Russell, 1 Normal Ave, Montclair, NJ 07043. E-mail: [russellm5@montclair.edu](mailto:russellm5@montclair.edu)

Abstract

Employee engagement has recently enjoyed a surge in interest as a positive employee outcome despite continued disageement regarding its factor structure and nomological relationship to other constructs, like burnout. We contrast two three-factor models of engagement: substantive, with the dimensions vigor, dedication and absorption, and attitudinal, with cognitive, affective and behavioral dimensions. Using bifactor analysis, study 1 proposes a scale that reconciles these two models and reduces 36 candidate items to 18. Study 2 convergently and discriminantly validates this scale.

*Keywords:* Engagement, bifactor analysis

*Word count:* X

Development of an Intentional BiFactor Engagement Measure

# 1 Introduction

The term engagement within the context of work emerged in organizational psychology and business literature in the early 90’s (@Kahn (1990)). Today, there are four main lines of research focused on engagement: personal engagement, burnout/engagement, work engagement, and employee engagement. These four constructs are defined and measured differently across the literature, making it hard to determine what exactly is being measured when organizations survey their employees to assess “engagement.” Work engagement has been defined as “the positive, fulfilling, work-related state of mind that is characterized by vigor, dedication, and absorption. Vigor is characterized by high levels of energy and mental resilience while working. Dedication refers to being strongly involved in one’s work and experiencing a sense of significance, enthusiasm, inspiration, pride, and challenge. Absorption is characterized by being fully concentrated and happily engrossed in one’s work, whereby time passes quickly and one has difficulties with detaching oneself from work” (Schaufeli et al. (2002)). Individuals who are engaged in their work are known to have high levels of energy, enthusiasm, and are completely immersed in their work tasks (Bakker et al., 2018). Based on the @Schaufeli et al. (2002) definition of engagement we conceptualize engagement as a mental state wherein employees feel energized, are enthusiastic about the content of their work and the things they do, and are so immersed in their work activities that time seems compressed.

There is a large body of evidence supporting the relationships between work engagement and organizational outcomes, including those which are performance based (Simpson, 2008). Which is why despite the lack of clarity around the construct, work engagement has emerged as an important component of an organization’s overall health and strategy. Given the potential outcomes of either having or lacking employees who experience work engagement. Across the literature you can find different trends in work engagement research, such as studying work engagement as a phenomenon that fluctuates within individuals, episodic work engagement, organizational-level work engagement, and leadership and work engagement. A relevant topic in recent literature is how dynamic work engagement can be. Bakker and colleagues (2018) point out that for organizational practice, it is important to understand that employees experience fluctuating levels of engagement at work and what causes these fluctuations (e.g., culture, environment, leadership, etc.). Although it is important that research continues to improve our understanding of the nature, causes, and outcomes of work engagement, it is also important that this knowledge is transferred to practical applications that can benefit both individuals and organizations. An important aspect of accumulating knowledge on the subject, is the methodological component of how that knowledge is built.

# 2 Models

Like many other constructs within the psychology literature, knowledge on work engagement has been built based on its measurement and linkage to other work outcomes. There are currently many work engagement scales used for either academic or applied purposes (sometimes both), which provide work engagement scores at the individual level. Most of these scales are measuring latent constructs that are not work engagement per se, but are intended to capture elements of this construct through indicators of other latent constructs. For example, the idea that vigor, dedication, and absorption together form the foundation of work engagement forms the basis of the Utrecht Work Engagement Scale (UWES; Schaufeli et al. (2002)). Despite its wide use and recognition, the UWES has been subject to criticism due to its development methodology and factorial structure (Willmer et al. (2019)). One of the main arguments being that the three subscales; vigor, dedication and absorption are closely correlated with each other, making it hard to argue that the three-factor structure is better than the single factor structure to measure work engagement (Kulikowski (2017a)). However, the UWES is still a widely used measure and its subscales are used to identify the specific subcategories of work engagement that need improvement. Given their popularity, the subscales within this theoretical framework are used as the first theoretical model for this project (see Figure 1). To our knowledge, the first use of the word “engagement” as a construct came from Kahn (1990), who defined it as: “the harnessing of organization members’ selves to their work roles; in engagement, people employ and express themselves physically, cognitively, and emotionally during role performances.” Although this definition was quickly bypassed by subsequent papers (see, for example, (Baumruk, 2004) and (Shaw, 2005), who framed it in terms of one’s cognitive and affective *commitment* to one’s organization), Kahn (1990)‘s definition is notable in that it conforms to the then-ascendant tripartite model of attitudes proposed by Rosenberg (1960). This model frames attitudes as latent variables that manifest cognitively, affectively and behaviorally. The tripartite model of attitude emerged from the Yale Communication and Attitude program in the 1950’s. It is a latent variable model based on the assumption that the latent variable elicits three types of responses or manifestations: a persons’ cognitive, affective, and behavioral responses to a specific attitude (Kaiser and Wilson (2019)). Even though it is not specifically a work engagement model, the tripartite model has helped researchers define and deconstruct attitudes to gain a better understanding of individuals’ responses towards specific attitude objects (Kaiser and Wilson (2019)).

# 3 Bifactor structures

Bifactor analysis is part of the exploratory factor analytic domain, and is used to account for variance in observed variables from the effects of latent or general factors (Giordano et al. (2020)). In most applications the latent constructs included in bifactor models are mutually orthogonal (i.e., uncorrelated), and represent the broad target constructs an instrument was designed to measure. Bifactor models are best suited to represent the multidimensionality arising from item responses that aim to measure broader constructs through multiple domains or subcategories (Reise (2012)). Additionally, these models can be classified as constrained hierarchical bifactor models and unconstrained non-hierarchical bifactor models (Giordano et al. (2020)). For the purposes of this project, an unconstrained non-hierarchical model is used to determine which group factors represent substantive and attitudinal constructs. The resulting general factors in the non-hierarchical model have a direct (i.e., non-mediated) effect on each variable observed. In this case, each general factor from both substantive and attitudinal models have non-mediated but overlapping effects on each variable.

we need to do some market research on the Q12: 1. what’s the feedback report look like? (google images show one overall “satsifaction” score and/or one overall “engagement” score), 2. how much does it cost, 3. what are the 200 pulse items Gallup refers to? (6/7/21)

This model is not without criticism, however. Some critics question its structural validity by pointing out that vigor, dedication and absorption all correlate highly with each other (Kulikowski, 2017b).

need more on criticisms of model

The present article explores two methods for constructing a scale that incorporates both the substantive and attitudinal models into one, a more classical one based on corrected item-total correlations and one based on modification indices.

# 4 Methods

Choice of focus on BIC versus AIC discussed in Dziak et al. (2020).

## 4.1 Participants

330 individuals provided ratings across 36 candidate items. These participants were gathered via snowball sampling, with an initial population of undergraduate and graduate students, as well as professional acquaintances of faculty members.` All surveys were administered on Qualtrics.

Participant job title, hours worked per week, and organizational tenure were recorded. Mean hours worked per week was 40.59 (SD = 13.69). Mean organizational tenure was 6.82 (SD = 8.50). Participants who did not exactly specify their tenure (e.g. “A bit over a year”) were not included in this average.

|  |  |
| --- | --- |
| Professional category | Count |
| Clerical Support Workers | 4 |
| Craft and related trades workers | 1 |
| Managers | 51 |
| Plant and machine operators, and assemblers | 3 |
| Professionals | 120 |
| Service and sales workers | 8 |
| Technicians and associate professionals | 62 |

Participants provided their job titles via an optional free text-entry box at the end of the survey. From there, we classified job titles according to the International Standard Classification of Occupations (ISCO-8) with the classify\_occupation function within the labourR package (Kouretsis et al. (2020)). The ISCO hierarchically organizes jobs in increasing order of specificity. For example, the first level of the hierarchy distinguishes a professional from a clerical worker or a technician. On the second level, professionals are distinguished among each other by whether they are engineers, medical workers, lawyers, and so on. See ??.

### 4.1.1 Item generation.

We generated a set of 50 items for our engagement measure, with the ultimate goal of reducing them to a final set of 18. These items were generated according to a review of extant tripartite engagement measures, as well as *WHAT RESEARCH DID WE USE FOR ATTITUDINAL WORDING? WAS IT LITERALLY JUST “I THINK,” “I FEEL,” “I DO?”* Each item was worded to reflect both a substantive dimension as well as an attitudinal dimension. For example, the item “My job makes me feel like I’m part of something meaningful” reflects the affective dimension with “feel” and the dedication dimension with “I’m part of something meaningful.”

Our 3x3 bifactor model produced nine pairs of dimensions (e.g., Vigor-Cognitive, Vigor-Affective, Vigor-Behavioral, etc.). With 36 initial items, this left four items per pair of substantive and attitudinal dimensions.

### 4.1.2 Content validation and initial item reduction.

An item sorting process was conducted to ensure content validity of our scale. Our original 50 items were presented to seven masters and PhD students in industrial-organizational psychology at Montclair State University, with each student instructed to sort each item into its respective substantive and attitudinal dimensions. Items that were not sorted into the same substantive-attitudinal dimension pair by at least five of seven raters were excluded from further analysis.

See table *X* for a full list of items and their respective dimensions.

Following item sorting, we further reviewed the wording of each item and eliminated all that, upon review, fell outside of the content domain (e.g. “I would rather work here than elsewhere”), eventually arriving at 36 candidate items.

## 4.2 Procedure

Looking into the specification of polychoric covariances (Jöreskog, 1994). This seems to be not very commonly leveraged (only package that seems to estimate these is semPlot).

The effective result of this was two divergent quasi-experimental approaches: 1) focus on corrected item-total correlations, and 2) focus on CFA modification indices.

### 4.2.1 Corrected item-total correlations.

To Casey: document your process here

We conducted a correct item-total correlation on our original 36-items set. Base off, the r. drops that the corrected item-total correlations provide us we narrowed it down by selecting that items that had the best r. drops off removing one item at a time. For example, each cell division contain 4 items, therefore, we remove one of the four items creating 6 potential 3 item corrected item correlations, and from there we choose the items with the best r. drops. We continued the same process when narrowing our three items down to two items. An example is shown below:

### 4.2.2 CFA Modification Indices.

We followed two parallel stepwise item-reduction processes centered around eliminating items in decreasing order of modification indices. Looking at the 36-item substantive and attitudinal models independently (process 1 and process 2), we requested modification indices from each, with the intent of retaining indicators whose fixed shared residual covariances were associated with high modification indices (indicating better model fit if the paths were freed). The item pair with the highest modification index was scrutinized, with a subjective group judgment made on wording and content domain coverage. The less preferred item was removed from the model. In cases where the highest modification index was between the only two remaining items in a substantive-attitudinal pair, these items were passed over for scrutiny in favor of the items with the next-highest index. This process was repeated until 18 items remained (i.e., 2 items for each of the 9 substantive-attitudinal pairs).

For example, the path with the highest modification index across both CFAs was between item 2 and item 4, which are both indicators of “Absorption” and “Cognition.” One of these items was therefore a candidate for deletion, and semantic preference was given to item 4, “I find it difficult to mentally disconnect from work” over item 2. After item 2 was excluded from both scale definitions (substantive and attitudinal), the CFAs were re-run and modification indices re-checked for bi-factor structure optimizing modifications.[[1]](#footnote-28)

The end result was two separate final scale definitions (one optimized for the substantive model and one for the attitudinal model).

Old text: We prioritized item deletions such that an item was implicated for deletion if: 1) modification index was high (relative to others) and 2) error residual was within the same “cell.” The choice of item to delete was based on author preference for wording/semantics as well as construct element coverage (considering the possible consequences for construct deficiency). Item variance was also consulted (retention more likely with greater item variance).

Table 1:

*Attitudinal structure modification indices (36 item analysis)*

|  |  |  |  |
| --- | --- | --- | --- |
| Element 1 | Element 2 | Modification Index | Notes |
| Item\_2 | Item\_4 | 192.41 | Candidate for deletion due to construct duplication |
| Item\_8 | Item\_18 | 96.05 |  |
| Item\_29 | Item\_35 | 62.25 | Candidate for retention due to substantive construct association |
| Item\_14 | Item\_20 | 56.38 |  |
| Item\_1 | Item\_12 | 51.39 |  |
| Item\_1 | Item\_13 | 50.33 |  |
| Item\_13 | Item\_12 | 41.40 |  |

Actually it doesn’t matter that much with only 1 item deletion - probably go ahead and do a few before recheck modification indices

### 4.2.3 Single factor versus bifactor approaches.

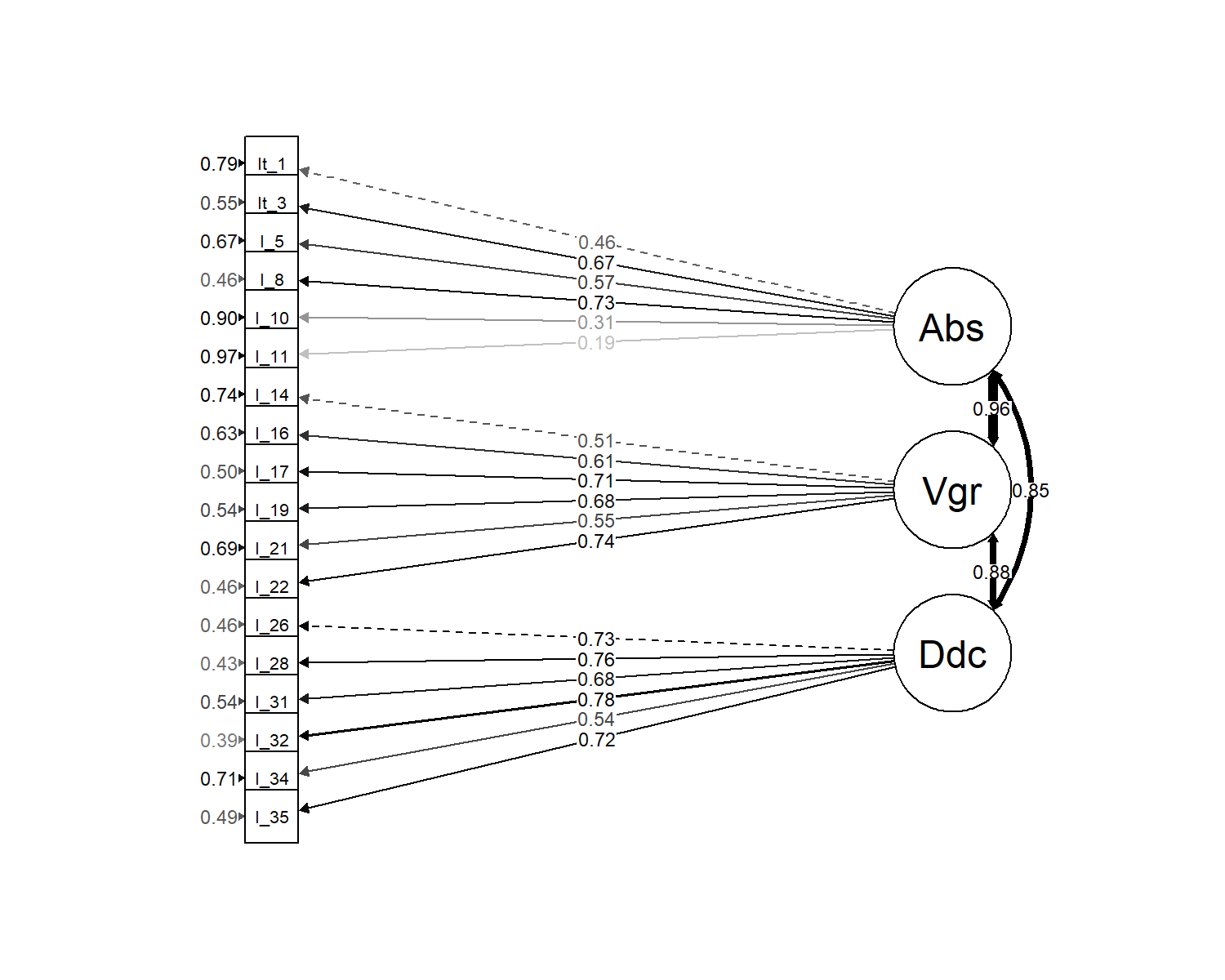
Casey this is where you come in

## 4.3 Data analysis

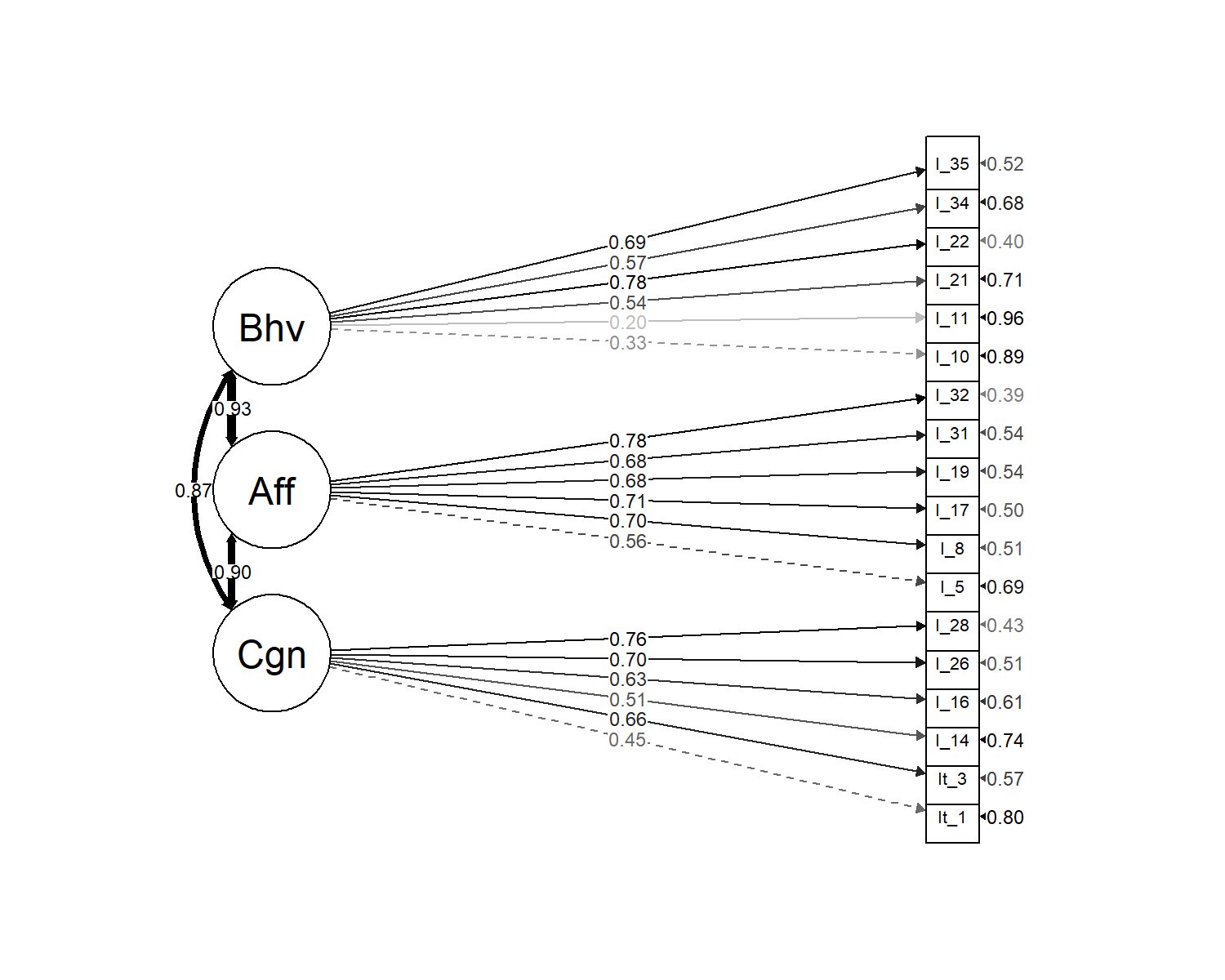
We used R [Version 4.1.0; R Core Team (2021)] and the R-packages *apaTables* [Version 2.0.8; Stanley (2021)], *dplyr* [Version 1.0.6; Wickham et al. (2021)], *DT* [Version 0.18; Xie et al. (2021)], *forcats* [Version 0.5.1; Wickham (2021a)], *ggplot2* [Version 3.3.3; Wickham (2016)], *kableExtra* [Version 1.3.4; Zhu (2021)], *labourR* [Version 1.0.0; Kouretsis et al. (2020)], *lavaan* [Version 0.6.8; Rosseel (2012)], *magrittr* [Version 2.0.1; Bache and Wickham (2020)], *papaja* [Version 0.1.0.9997; Aust and Barth (2020)], *purrr* [Version 0.3.4; Henry and Wickham (2020)], *readr* [Version 1.4.0; Wickham and Hester (2020)], *sem* [Version 3.1.11; Fox et al. (2020); Epskamp (2019)], *semPlot* [Version 1.1.2; Epskamp (2019)], *stringr* [Version 1.4.0; Wickham (2019)], *tibble* [Version 3.1.2; Müller and Wickham (2021)], *tidyr* [Version 1.1.3; Wickham (2021b)], and *tidyverse* [Version 1.3.1; Wickham et al. (2019)] for all analyses.

# 5 Results

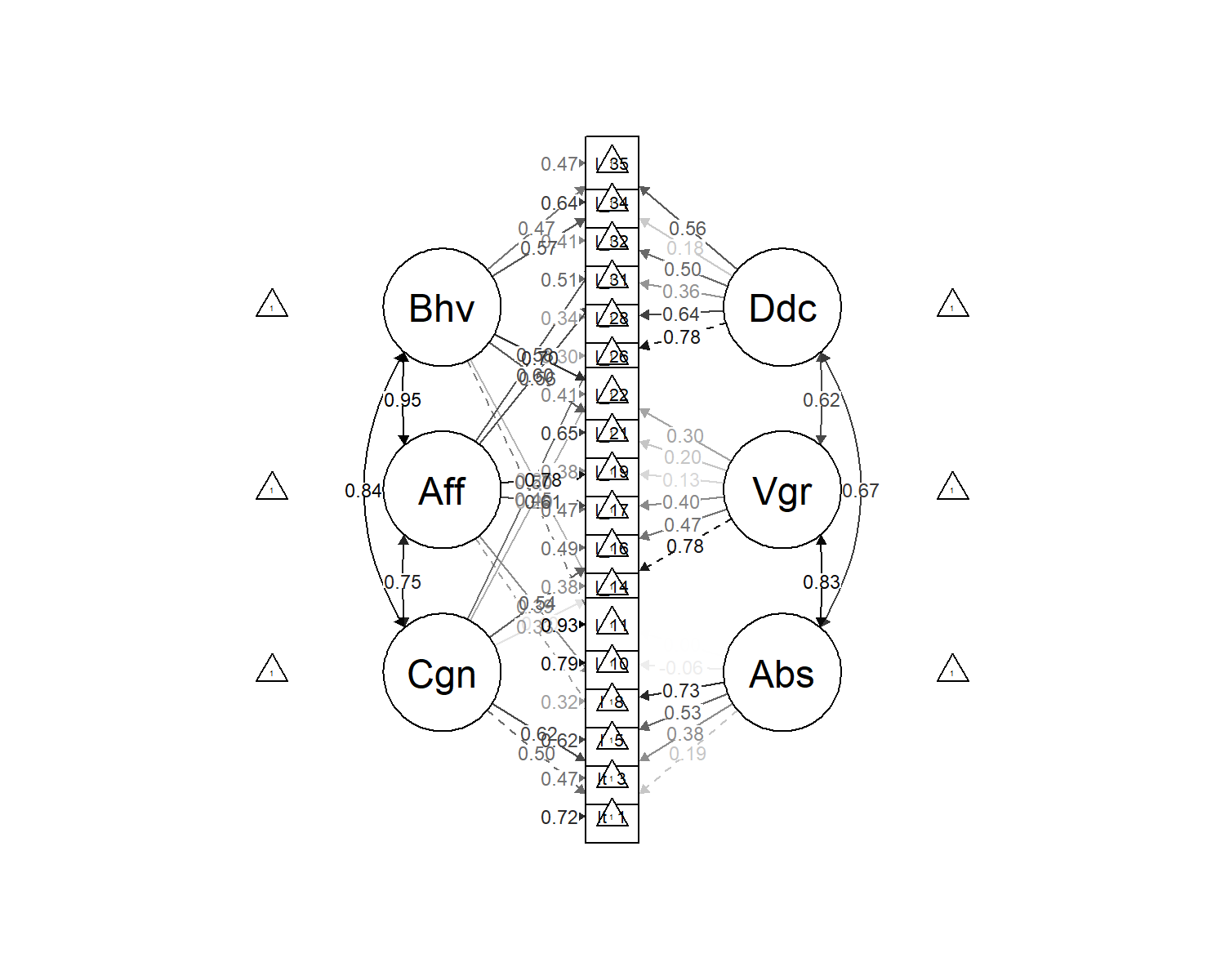
CFA drafts below



*Figure* *1.*  Substantive factor measurement model



*Figure* *2.*  Attitudinal factor measurement model



*Figure* *3.*  Bifactor measurement model

Table 2:

*Summary fit statistics (18 item proposed scale definitions)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| model | Chi Square | df | RMSEA | SRMR | CFI | TLI | AIC |
| Attitudinal | 454.30 | 132.00 | 0.10 | 0.07 | 0.83 | 0.80 | 13,473.40 |
| Substantive | 473.56 | 132.00 | 0.10 | 0.07 | 0.82 | 0.79 | 13,492.66 |
| bifactor | 264.70 | 111.00 | 0.07 | 0.05 | 0.92 | 0.89 | 14,113.31 |

## 5.1 Final scale definitions

## 5.2 Study 2

Construct validation was acccomplished via administration of the 17-item UWES as well as the Saks (2006) 12-item scale. Saks (2006) aggregates to two scales: job and organizational engagement.

# 6 Discussion

The purpose of this study was to present two divergent approaches for constructing scales that simultaneously probe the substantive and attitudinal factor structures of employee engagement. Toward this end, we propose two similar scale definitions

Our next endeavor will be to establish convergent and discriminant validity of the scales.

Currently, the UWES is one of the most popular engagement scales Our proposed scale would

Bifactor analysis may be a valuable method of reconciling previously at-odds approaches to describing the same construct.

Our research contributes to theory in two key ways. Firstly, it introduces a novel measure of engagement, developed in English, that will allow future researchers to further probe the tripartite attitudinal structure of the construct. To our knowledge, ours is the only engagement scale that probes the specific attitudinal dimensions of engagement. Secondly, we more generally advance the use of bifactor analysis as an alternative approach to testing and comparing structural models of constructs. Rather than We show that a scale can exhibit high internal consistency while simultaneously measuring two different structural models. It is our hope that the success of this approach might evangelize bifactor analysis and the more general approach of consolidating and integrating theories rather than parsing them.

# 7 References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. <https://github.com/crsh/papaja>

Bache, S. M., & Wickham, H. (2020). *Magrittr: A forward-pipe operator for r*. <https://CRAN.R-project.org/package=magrittr>

Baumruk, R. (2004). *The missing link: The role of employee engagement in business success*. *47*, 48–52.

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2020). Sensitivity and specificity of information criteria. *Briefings in Bioinformatics*, *21*(2), 553–565.

Epskamp, S. (2019). *semPlot: Path diagrams and visual analysis of various SEM packages’ output*. <https://CRAN.R-project.org/package=semPlot>

Fox, J., Nie, Z., & Byrnes, J. (2020). *Sem: Structural equation models*. <https://CRAN.R-project.org/package=sem>

Giordano, C., Ones, D. S., Waller, N. G., & Stanek, K. C. (2020). Exploratory bifactor measurement models in vocational behavior research. *Journal of Vocational Behavior*, *120*, 103430. <https://doi.org/10.1016/j.jvb.2020.103430>

Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools*. <https://CRAN.R-project.org/package=purrr>

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, *59*(3), 381–389.

Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, *33*(4), 692–724.

Kaiser, F. G., & Wilson, M. (2019). The campbell paradigm as a behavior-predictive reinterpretation of the classical tripartite model of attitudes. *European Psychologist*, *24*(4), 359–374. <https://doi.org/10.1027/1016-9040/a000364>

Kaiser, F. G., & Wilson, M. (2019). The campbell paradigm as a behavior-predictive reinterpretation of the classical tripartite model of attitudes. *European Psychologist*, *24*(4), 359–374. <https://doi.org/10.1027/1016-9040/a000364>

Kouretsis, A., Bampouris, A., Morfiris, P., & Papageorgiou, K. (2020). *labourR: Classify multilingual labour market free-text to standardized hierarchical occupations*. <https://CRAN.R-project.org/package=labourR>

Kulikowski, K. (2017a). Do we all agree on how to measure work engagement? Factorial validity of utrecht work engagement scale as a standard measurement tool - a literature review. *International Journal of Occupational Medicine and Environmental Health*, *30*(2), 161–175. <https://doi.org/10.13075/ijomeh.1896.00947>

Kulikowski, K. (2017b). Do we all agree on how to measure work engagement? Factorial validity of utrecht work engagement scale as a standard measurement tool–a literature review. *International Journal of Occupational Medicine and Environmental Health*, *30*(2), 161–175.

Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames*. <https://CRAN.R-project.org/package=tibble>

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. <https://doi.org/10.1080/00273171.2012.715555>

Rosenberg, M. J. (1960). Cognitive, affective, and behavioral components of attitudes. In *Attitude organization and change*.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. <https://www.jstatsoft.org/v48/i02/>

Saks, A. M. (2006). Antecedents and consequences of employee engagement. *Journal of Managerial Psychology*, *21*(7), 600–619.

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

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

Shaw, K. (2005). An engagement strategy process for communicators. *Strategic Communication Management*, *9*(3), 26.

Stanley, D. (2021). *apaTables: Create american psychological association (APA) style tables*. <https://CRAN.R-project.org/package=apaTables>

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>

Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations*. <https://CRAN.R-project.org/package=stringr>

Wickham, H. (2021a). *Forcats: Tools for working with categorical variables (factors)*. <https://CRAN.R-project.org/package=forcats>

Wickham, H. (2021b). *Tidyr: Tidy messy data*. <https://CRAN.R-project.org/package=tidyr>

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. <https://doi.org/10.21105/joss.01686>

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of data manipulation*. <https://CRAN.R-project.org/package=dplyr>

Wickham, H., & Hester, J. (2020). *Readr: Read rectangular text data*. <https://CRAN.R-project.org/package=readr>

Willmer, M., Westerberg Jacobson, J., & Lindberg, M. (2019). Exploratory and confirmatory factor analysis of the 9-item utrecht work engagement scale in a multi-occupational female sample: A cross-sectional study. *Frontiers in Psychology*, *10*. <https://doi.org/10.3389/fpsyg.2019.02771>

Xie, Y., Cheng, J., & Tan, X. (2021). *DT: A wrapper of the JavaScript library ’DataTables’*. <https://CRAN.R-project.org/package=DT>

Zhu, H. (2021). *kableExtra: Construct complex table with ’kable’ and pipe syntax*. <https://CRAN.R-project.org/package=kableExtra>

1. Probably put a table in here highlighting certain modification indices (with a key to intended factor-item association). Look at “modincides1” [↑](#footnote-ref-28)