Contents lists available at Vilnius University Press

The Reliability of Positive and Negative Work Reflection Scales: Reliability Generalization Meta-Analysis

Tadas Vadvilavičius

Vytautas Magnus University
tadas.vadvilavicius@vdu.lt
https://orcid.org/0000-0002-1920-1959
https://ror.org/04y7eh037

Abstract. Positive and Negative work reflection are important constructs in the field of organizational psychology. To measure these constructs, reliable measures are needed. In order to summarize the findings from different studies, a reliability generalization meta-analysis was performed. A literature search in six databases was conducted. In total, 216 records were found, while 26 papers were included in the analysis. Positive and Negative work reflection scales by Fritz and Sonnentag (2005) and Binnewies et al. (2009) were analyzed. Analysis revealed that internal consistency varied from 0.90 to 0.93. Additionally, the region in which a study was conducted moderated part of the results.

Keywords: meta-analysis, positive work reflection, negative work reflection, reliability.

Pozityvios ir negatyvios darbo refleksijos skalių vidinis suderinamumas: patikimumo apibendrinimo metaanalizė

Santrauka. Pozityvi ir negatyvi darbo refleksija organizacinės psichologijos srityje yra svarbūs konstruktai. Norint išmatuoti šiuos konstruktus, reikia patikimų matavimo priemonių. Siekiant apibendrinti skirtingų tyrimų rezultatus, buvo atlikta patikimumo apibendrinimo metaanalizė. Atlikta literatūros paieška šešiose duomenų bazėse. Iš viso rasta 216 įrašų, o į analizę įtraukti 26 straipsniai. Buvo išanalizuotos Fritz ir Sonnentag (2005) ir Binnewies ir kt. (2009) pozityvios ir neigiamos darbo refleksijos skalės. Analizė parodė, kad vidinis suderinamumas svyravo nuo 0,90 iki 0,93. Be to, regionas, kuriame buvo atliktas tyrimas, turėjo įtakos daliai rezultatų.

Pagrindiniai žodžiai: metaanalizė, pozityvi darbo refleksija, negatyvi darbo refleksija, patikimumas.

Introduction

Thinking about work is an important and interesting topic in the field of organizational psychology. However, as it has been discussed elsewhere, previous research focused more on rumination, the inability to cognitively switch from work during non-work time, or low psychological detachment from work (Binnewies et al., 2009). Fritz and Sonnentag (2005) have discussed that work reflection is an important mental process which occurs during

leisure time and may be related to various outcomes for organizations and employees, like recovery from work.

Work reflection can be both positive and negative, and refers to thinking positively or negatively about work after work hours (Binnewies et al., 2009; Fritz & Sonnentag, 2005, 2006). Studies have shown that positive work reflection is related to higher work engagement (e.g., Ilies et al., 2024) and work-family enrichment (e.g., Kim & Beehr, 2023), lower burnout (e.g., Tong & Spitzmueller, 2024), etc. Meanwhile, negative work reflection is related to lower work engagement (e.g., Ilies et al., 2024), higher depression, and angry mood (e.g., Meier et al., 2016). Positive work reflection may help employees cope with work-related stress and increase work meaningfulness, while negative work reflection may prolong stress (e.g., Haun & Oppenauer, 2019; Sonnentag et al., 2021). In general, both positive and negative work reflection allows to predict organizational and personal outcomes, like task performance, creativity, strain (e.g., Binnewies et al., 2009; Xu et al., 2021). This important construct requires reliable tools for its measurement. It is of importance to emphasize that a deeper theoretical investigation of the relationship between Positive and Negative work reflection is needed, as there are studies showing that these two variables are not inter-related (e.g., Walter & Haun, 2020).

Two scales were developed to measure work reflection, as each scale has a version for both positive and negative work reflection (thus producing four scales in total). The first scale was developed by Fritz and Sonnentag (2005) and consisted of three items, while, a few years later, Binnewies et al. (2009) presented a modified scale with one additional item. Both scales were used in various countries, e.g., Germany, China, and the U.S.; however, the psychometric properties of the scales are still rarely analyzed. This study focuses solely on the internal consistency of the scales. Reliability generalization meta-analysis is used for evaluating the consistency of the measurement instruments across studies. A meta-analysis can offer a robust and precise estimation of the reliability of both Positive and Negative work reflection scales across studies. This helps to understand the average and variability of reliability coefficients of both scales, and that the high (if these are truly high) internal consistencies are not a coincidence. The main research question of this research is: *What is the overall internal reliability of the positive and negative work reflection scales?*

Methods

Literature search

Six databases were reviewed: *EBSCO Academic Ultimate*, *ScienceDirect*, *Scopus*, *Web of Science*, *Emerald Management eJournals Collection*, and *SpringerJournals* collection. These databases were chosen based on the quality of databases, accessibility, and coverage. Four keywords were used: "*positive work reflection*", "*positive reflection at work*", "*negative work reflection*", and "*negative reflection at work*". Keywords were used to search for relevant literature by both titles and abstracts. The search procedure was not

restricted by the date or the region where the research was conducted. Search algorithms and results can be found in the Supplementary material page on the OSF website.

Three inclusion criteria were set *prior to* the analysis: 1) primary data was used and presented on a paper; 2) positive work reflection was measured; 3) a paper reported internal consistency by using Cronbach's alpha or McDonald's Omega. Exclusion criteria were also set *prior to* the analysis: 1) a qualitative study design, or other article types, not presenting primary quantitative results (e.g., opinion papers, commentaries); 2) secondary studies (e.g., meta-analysis); 3) positive work reflection was not measured; 4) internal

Figure 1

Flowchart of the process of the selection of studies for the present research
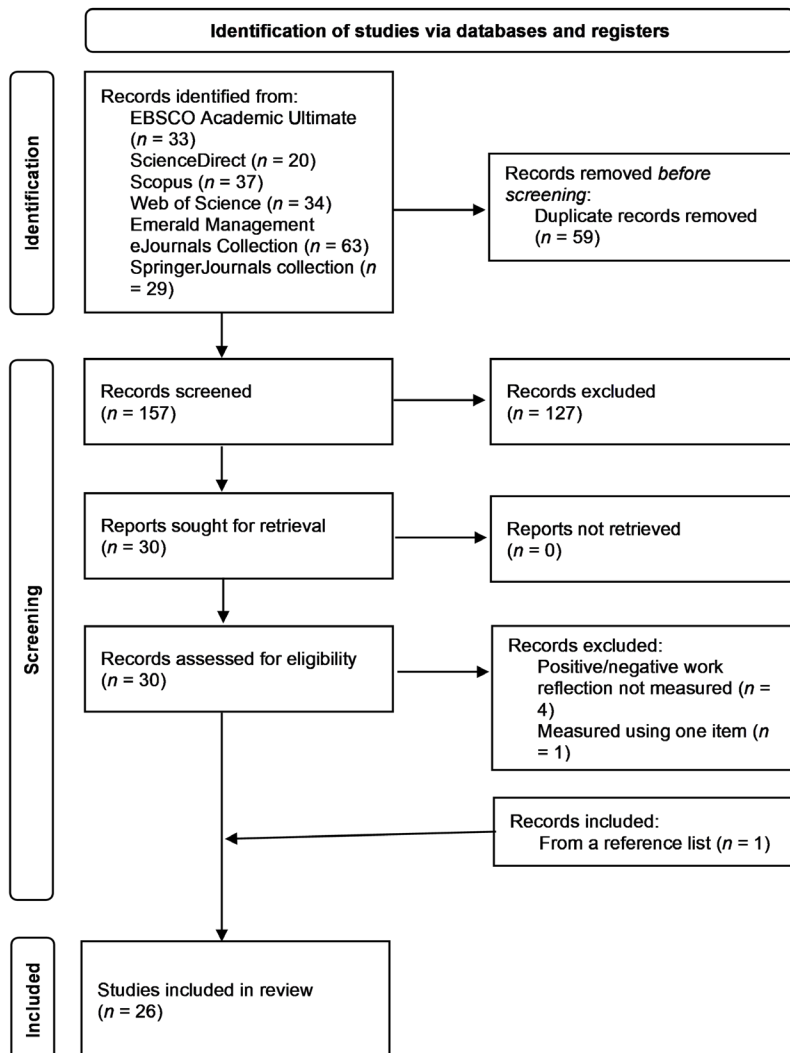
consistency was not presented. Additionally, a review of reference lists of selected articles was done in order to identify potential articles based solely on their titles. This additional step allowed to identify one extra article. The final list of papers included into analysis is provided in the Supplementary material section.

Rayyan.ai software was used for the screening process. At first, only the titles and abstracts were screened. Figure 1 presents the flow of the review process. Data from the studies were extracted manually. The data extracted were coded in an SPSS file; the data cover the authors, the country where research was conducted, the number of participants, and the obtained Cronbach's alpha coefficients. If a study was longitudinal and provided few effect sizes, only one from the previous measure point was used, e.g., only T1, if provided. For diary studies, effect sizes for the within-level or mean across days, if provided, were used. If not, the lowest score was used. The data file for SPSS is provided in the Supplementary material section.

To enhance the data extraction reliability and given that the review was conducted by a single researcher, the extracted data and the coded SPSS file were independently re-verified a few days after completing the initial extraction. One effect size was missing, and one typo was made. Quality assessment of the studies was not conducted. The study was not pre-registered.

Meta-analysis

Meta-analysis was performed by using the *metafor* (Viechtbauer, 2010) package for *R* (R Core Team, 2024) and *RStudio* (RStudio Team, 2025). Heterogeneity between studies was assessed by using Q and I^2 statistics. A significant Q score indicates heterogeneity between effects, whereas I^2 indicates the percentage of between-effect variance that is not a sampling error. A higher I^2 statistic represents higher heterogeneity. The funnel plot and Egger's-test were used to test publication bias. The level of statistical significance was set at $p < .05$ (two-sided). For more details about the statistical analysis, see Edelsbrunner et al. (2025).

A random-effects model was used because it cannot be assumed that all studies are from a single population. Firstly, Cronbach's alpha and McDonald's Omega coefficients were transformed by using a logit-transformation. After calculating and pooling logit(alpha) results, the results were transformed back into Cronbach's alpha/McDonald's Omega.

Results

In total, 216 records were found from all six databases. The full review of the articles under consideration revealed that 25 articles meet the inclusion criteria. An additional article was identified by the reference list review, resulting in a total of 26 articles. These 26 articles presented 30 studies in total. Most of the studies were conducted in Germany ($n = 10$), followed by those conducted in the United States ($n = 8$) and China ($n = 4$). Meanwhile, most of the studies were published in the period between 2019 and 2024 ($n = 19$). Finally,

the sample size ranged from 59 to 1036 participants, with a mean number of participants at 227.10 (SD = 192.88).

From the 26 papers reviewed, 14 papers reported using the positive work reflection scale by Fritz & Sonnentag (2005), nine reported using the negative work reflection scale by Fritz & Sonnentag (2005), six reported using the positive work reflection scale by Binnewies et al. (2009), and seven reported using the negative work reflection scale by Binnewies et al. (2009).

The results revealed that the internal reliability was in all cases above .90 (see Table 1), revealing a high internal reliability for all scales. Meanwhile, the use of the random-effects model was confirmed by significant Q statistics and high I^2 scores. Forrest plot can be found in the Supplementary material. section

Table 1

Effect-size summary statistics

Authors	Scale	No. of effects	Total sample size	Combined Cronbach's alpha coefficients (95% CI)	Heterogeneity test	I^2 (%; 95% CI)
Fritz & Sonnentag (2005)	Positive work reflection scale	16	4077	.90 (.85; .93)	$Q(15) = 556.13$, $p < .001$	97.52 (95.44; 98.97)
	Negative work reflection scale	9	1835	.90 (.82; .95)	$Q(8) = 356.51$, $p < .001$	98.07 (95.74; 99.48)
Binnewies et al. (2009)	Positive work reflection scale	7	1812	.90 (.85; .94)	$Q(6) = 82.33$, $p < .001$	95.75 (89.16; 99.18)
	Negative work reflection scale	7	2093	.93 (.91; .94)	$Q(6) = 48.63$, $p < .001$	89.65 (73.07; 97.85)

Additionally, moderation analysis was performed to test whether the region in which the study was conducted affected the results (see Table 2). Analysis revealed that Cronbach's alpha coefficients are statistically significantly higher for both Positive and Negative work reflection scales by Fritz and Sonnentag (2005) in the U.S., compared to European or Asian research.

Finally, Egger's test revealed no publication bias for the Fritz and Sonnentag (2005) Positive work reflection ($z = .42$, $p = .67$), the Fritz and Sonnentag (2005) Negative work reflection scale ($z = 1.86$, $p = .06$), and the Binnewies et al. (2009) Positive work reflection ($z = .52$, $p = .61$). Meanwhile, publication bias was observed when analyzing the Binnewies et al. (2009) Negative work reflection scale ($z = 2.57$, $p < .05$); however, the results can be affected by the low number of the effect sizes used. Additionally, Fail-safe

N test (or file drawer analysis) revealed that 322 ($p < .001$) effect sizes should be included in the meta-analysis for the effect size to be insignificant, thereby revealing the robustness of the results for the Binnewies et al. (2009) Negative work reflection scale.

Table 2

Moderation analysis

Authors	Scale	No. of effects	Test of moderator	Country estimates	I^2 (%)
Fritz & Son-nentag (2005)	Positive work reflection scale	16	$QM(2) = 23.48$, $p < .001$	North America .95	93.79
				Europe .90	
				Asia .81	
	Negative work reflection scale	9	$QM(2) = 22.68$, $p < .001$	North America .98	92.52
				Europe .88	
				Asia .81	
Binnewies et al. (2009)	Positive work reflection scale	7	$QM(1) = 1.27$, $p = .26$	North America .94	95.75
				Europe .89	
				NA	
	Negative work reflection scale	7	$QM(2) = 0.60$, $p = .43$	North America .95	90.47
				Europe .94	
				NA	

Discussion

This reliability generalization meta-analysis assessed the internal consistency of Positive and Negative work reflection scales. The results revealed high internal consistency for both the three-item Positive and Negative work reflection scales of Fritz and Sonnentag (2005) and the four-item Binnewies et al. (2009) scales.

The final number of papers used in the review was 26 papers. However, the search and analysis are not without limitation. Studies only in the English language were found and analyzed. There is at least one known study (Wang et al., 2024) which does present the necessary data for the analysis, however, as it was not identified during the search procedure, it was consequently *not* included. It can be assumed that the paper was published in a journal that was not referred in the analyzed databases. Finally, only two strategies to find the necessary body of literature was used. In the future, more strategies could be used, for example, one could look for gray literature in *Google Scholar*, reviewing studies citing the articles identified, etc.

Most of the studies were published in Germany, which is not surprising as the authors of both scales represent Germany. It can be assumed that the researchers are leading authors in the field. Meanwhile, most of the studies were published in the period between 2019 and 2024, suggesting the rise of interest in this topic. However, the interest in the topic is without a clear reason. The sample size ranged from 59 to 1036 participants, which is

suitable, considering the research designs used, e.g., a diary study, or a longitudinal study. Additionally, it is of importance to highlight that quality assessment was not performed, which may result in biased results. Finally, the results may be affected due to expected data slicing in a few articles presented by the same authors from similar samples.

The positive work reflection scale by Fritz and Sonnentag (2005) was used most commonly. The results revealed a high internal consistency of .90. Moderation analysis revealed that the internal consistency of the scale is the highest in the U.S., compared to Europe and Asia. The negative work reflection scale by Fritz and Sonnentag (2005) was used more frequently compared to the scale by Binnewies et al. (2009). The internal consistency of the Negative work reflection scale was the highest in the U.S., compared to Europe and Asia. Possible language and cultural differences should be tested in the future. Things like word ambiguity, inconsistent translations, and the cultural meaning of words can lead to differences across countries and, consequently, to a lower internal consistency. Therefore, invariance analysis should be conducted. Additionally, the internal consistency should be tested in different samples, for example, based on professions, age, etc. Finally, both the Positive work reflection scale and the Negative work reflection scale by Binnewies et al. (2009) revealed a high internal consistency of 0.90 and 0.93, respectively. Moderation analysis revealed no differences between different regions of the world.

Positive/Negative work reflection is unidimensional, and therefore no factor-level analysis was conducted. Furthermore, while it is discussed that a larger number of items in a scale will result in higher internal consistency and less prominent measurement error (e.g., Böckenholt & Lehmann, 2015; Robinson, 2018), the internal reliability of the four-item Binnewies et al. (2009) scale was revealed to be almost the same as for the Fritz and Sonnentag (2005) scale (considering that internal consistency for the Positive work reflection scales is the same, and the internal consistency for the Negative work reflection scale differs only by .03). Thus, it could be discussed that both versions (i.e., both three- and four-item scales) are reliable and can be used in future research. Additionally, Xu et al. (2018) used a single item from the Fritz and Sonnentag (2005) Positive work reflection scale. It could be discussed that, because Positive work reflection is a unidimensional construct, a single item could be used for measuring this construct (e.g., Böckenholt & Lehmann, 2015). However, further studies are still needed regarding this specific point.

Supplementary Material

The supplementary material can be found in Open Science Framework website: https://osf.io/nzkwr/?view_only=1501ca7390044d08b2b9442c417e663b

References

Binnewies, C., Sonnentag, S., & Mojza, E. J. (2009). Feeling recovered and thinking about the good sides of one's work. *Journal of Occupational Health Psychology, 14*(3), 243–256. https://doi.org/10.1037/a0014933

Böckenholt, U., & Lehmann, D. R. (2015). On the limits of research rigidity: the number of items in a scale. *Marketing Letters*, *26*(3), 257–260. https://doi.org/10.1007/s11002-015-9373-y

Edelsbrunner, P. A., Simonsmeier, B. A., & Schneider, M. (2025). The Cronbach's alpha of domain-specific knowledge tests before and after learning: A meta-analysis of published studies. *Educational Psychology Review*, *37*(1), Article 4. https://doi.org/10.1007/s10648-024-09982-y

Fritz, C., & Sonnentag, S. (2005). Recovery, Health, and Job Performance: Effects of Weekend Experiences. *Journal of Occupational Health Psychology, 10*(3), 187–199. https://doi.org/10.1037/1076-8998.10.3.187

Fritz, C., & Sonnentag, S. (2006). Recovery, well-being, and performance-related outcomes: The role of workload and vacation experiences. *Journal of Applied Psychology, 91*(4), 936–945. https://doi.org/10.1037/0021-9010.91.4.936

Haun, V. C., & Oppenauer, V. (2019). The role of job demands and negative work reflection in employees' trajectory of sleep quality over the workweek. *Journal of Occupational Health Psychology, 24*(6), 675–688. https://doi.org/10.1037/ocp0000156

Ilies, R., Liu, Y., Aw, S., Las Heras, M., & Rofcanin, Y. (2024). Why does using personal strengths at work increase employee engagement, who makes the most out of it, and how?. *Journal of Occupational Health Psychology*, *29*(2), 113–129. https://doi.org/10.1037/ocp0000374

Kim, M., & Beehr, T. A. (2022). Can reflection explain how empowering leadership affects spillover to family life? Let me think about it. *The International Journal of Human Resource Management.* Advance online publication. https://doi.org/10.1080/09585192.2022.2054282

Meier, L. L., Cho, E., & Dumani, S. (2016). The effect of positive work reflection during leisure time on affective well-being: Results from three diary studies. *Journal of Organizational Behavior*, *37*(2), 255–278. https://doi.org/10.1002/job.2039

RStudio Team. (2025). *RStudio: Integrated development for R*. RStudio. http://www.rstudio.com/

R Core Team. (2025). *R: A language and environment for statistical computing* (Version 2025.05.0+496). R Foundation for Statistical Computing, Austria. https://www.R-project.org/

Robinson, M. A. (2018). Using multi-item psychometric scales for research and practice in human resource management. *Human Resource Management*, *57*(3), 739–750. https://doi.org/10.1002/hrm.21852

Sonnentag, S., Tian, A. W., Cao, J., & Grushina, S. V. (2021). Positive work reflection during the evening and next-day work engagement: Testing mediating mechanisms and cyclical processes. *Journal of Occupational and Organizational Psychology*, *94*(4), 836–865. https://doi.org/10.1111/joop.12362

Tong, J., & Spitzmueller, C. (2024). To reflect or to detach after work? Relating interpersonal challenge and hindrance stressors to engagement and burnout. *International Journal of Stress Management*, *31*(2), 184–195. https://doi.org/10.1037/str0000317

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of statistical software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Wang, Z., Meng, L., Cai, S., & Jiang, L. A. (2024). Work reflection during leisure time and employee creativity: The role of psychological capital. *Journal of Management & Organization*, *30*(2), 318–330. https://doi.org/10.1017/jmo.2020.10

Walter, J., & Haun, V. C. (2021). Positive and negative work reflection, engagement and exhaustion in dual-earner couples: Exploring living with children and work-linkage as moderators. *German Journal of Human Resource Management: Zeitschrift für Personalforschung*, *35*(2), 249–273. https://doi.org/10.1177/2397002220964930

Xu, S., Martinez, L. R., Van Hoof, H., Estrella Duran, M., Maldonado Perez, G., & Gavilanes, J. (2018). Emotional exhaustion among hotel employees: The interactive effects of affective dispositions and positive work reflection. *Cornell Hospitality Quarterly*, *59*(3), 285–295. https://doi.org/10.1177/1938965517748774

Xu, X., Jiang, L., Hong, P. Y., & Roche, M. (2021). Will mindful employees benefit from positive work reflection triggered by transformational leadership? A two-study examination. *International Journal of Stress Management*, *28*(1), 61–73. https://doi.org/10.1037/str0000222