merits

Brief Report

Does Changing a Scale's Context Impact Its Psychometric Properties? A Comparison Using the PERMA-Profiler and the Workplace PERMA-Profiler

Sean P. M. Rice [1,2] (ID)

1   School of Public Health, Oregon Health & Science University—Portland State University, Portland, OR 97201, USA; ricese@ohsu.edu
2   Oregon Institute of Occupational Health Sciences, Oregon Health & Science University, Portland, OR 97239, USA

Abstract: The present study evaluated the empirical distinction between the PERMA-Profiler and the Workplace PERMA-Profiler, which measure flourishing using the same items with different contexts (i.e., general vs. workplace orientations). Both scales were administered online via MTurk (N = 601), and single-group measurement and structural invariances were assessed. Partial metric and scalar invariances were supported, indicating that the PERMA constructs were measured equivalently across scales (except for the relationships factor). Structural properties (covariances, means) were not invariant, indicating distinct utility for each scale in their respective contexts. The results suggest that simple adaptations to items to change their context, but not content, may retain the original scale's psychometric properties and function with discrimination.

Keywords: psychometrics; flourishing; measurement invariance; scale adaptation; positive psychology

1. Introduction

Scales are often adapted for the context in which they are used, which may involve simply adjusting items to orient the respondent to the study's focal context. However, if a scale's items are adapted to be context-specific (e.g., moving from a general context to a workplace context), will its psychometric properties change? This study aimed to investigate this question by evaluating the measurement and structural equivalence of the PERMA-Profiler (PP) [1] and its workplace adaptation, the Workplace PERMA-Profiler (WPP) [2].

One conceptualization of wellbeing has been gaining traction in the last decade: Seligman's PERMA model of flourishing [3]. According to Seligman, flourishing is a higher-order wellbeing concept comprising positive emotions (P), engagement (E), relationships (R), meaning (M), and accomplishment (A). The PP [1], measuring one's general flourishing, and the WPP [2], measuring one's flourishing at work, have been developed to directly measure this type of flourishing. Although some argue that general wellbeing and work wellbeing are fundamentally different [4], the WPP is a simple adjustment to the items of the PP, making them workplace-specific (i.e., same construct, different context).

One of the reasons the PP and the WPP were selected for the present study was due to the fact that both versions have been previously validated and used in published research, which is rare for simple item context adjustments (but see Luthans et al.'s [5] and Lorenz et al.'s [6] psychological capital questionnaires for other examples). As indicated above, in many cases, when an original scale does not quite fit the context or population, slight adjustments to item wording are implemented (e.g., Culbertson et al.'s [7] workplace adaptation of Ryff's [8] psychological wellbeing scale). Additionally, the workplace adaptation has been translated and validated in other languages (e.g., Japanese [9] and Korean [10]), with studies identifying statistical distinctions between the WPP and PP (though without

direct comparisons). Finally, more recent research has summarily evaluated the psychometric properties of PP and WPP across a number of samples, concluding that there may be problems with the WPP's item wordings [11,12]. Such issues pose a problem for utilization of the WPP in workplaces (its intended context), as leaders wanting to measure their employees' (or their own) flourishing at work may receive questionable data. As such, continued evaluation of the adaptation and statistical comparisons with its parent are necessary if this scale is to be continually used.

In some cases, contextual scale adaptations are used in studies without any statistical evaluation of their consistency with the original. For example, in order to measure happiness in children, Holder and Klassen [13] slightly reworded Lyubomirsky and Lepper's Subjective Happiness Scale [14] for a younger population (e.g., changing "To what extent does this characterization describe you?" to "How much does this sentence describe you?"). Although seemingly innocuous, it is important to note that Holder and Klassen's version of the scale had poor internal consistency (Cronbach's alpha = 0.67, which is lower compared to adult applications of the scale [14,15]) [13]. Although the authors found poor reliability, and no work was done to either validate the adaptation or evaluate it for consistency with the original, this adaptation was used in future studies (e.g., Layous et al. [16], who notably did not report any psychometric properties), which calls into question whether the findings themselves may be reliable. For instance, it may not be that Layous et al. kindness intervention was ineffective at improving happiness in the preadolescents [16], but rather the measure they used did not function appropriately. As such, contextual adaptations to a scale that do not necessitate content adjustments may have impactful influences, not just on the measurement of a construct, but on the efficacy of interventions.

In a systematic review of workplace interventions to reduce mental health stigma, Tóth et al. [17] found that general stigma assessments may not have been sensitive enough to detect changes in some interventions due to their lack of a workplace-context focus. Similarly, a meta-analysis of Acceptance and Commitment Therapy (ACT) interventions to reduce dysregulated eating behaviors (e.g., binge eating) found that general psychological flexibility scales (e.g., the Acceptance and Action Questionnaire) were associated with significantly smaller RCT effects compared to their weight-specific adaptations (e.g., the Acceptance and Action Questionnaire for Weight-Related Difficulties) [18]. In summary, it may be necessary for individuals applying interventions in a particular context (e.g., managers implementing a new intervention to improve their employees' wellbeing) to adapt a "general" scale for said context (e.g., the workplace) in order to know whether the intervention was effective. However, without appropriately evaluating and understanding how those adaptations can change the scale's function and validity, it may cause idiosyncrasies across implementations and ultimately produce unreliable results.

In the present study, I aimed to empirically illustrate what impacts changing the context or wording of a scale's items may have on its psychometric properties using two previously validated scales of flourishing. Because the PP and the WPP were both developed to measure the same constructs in different contexts, I expected their respective measurement components (e.g., item loadings) to be equivalent. The structural components (e.g., inter-factor covariances, like-factor means), however, may or may not differ by context (e.g., a sample's level of perceived life meaning could be higher or lower or the same as their perceived work meaning). Our work was guided by the following research questions:

RQ1. Are the measurement properties of the PP and the WPP invariant? Specifically, are the patterns of latent PERMA constructs equivalent across scales (configural invariance), are the associations between the latent constructs and their respective like-items equivalent (metric invariance), and does each construct, when at an equal level, produce equivalent averages across like-items (scalar invariance)?

RQ2. Are the structural properties (i.e., factor variances and means) of PERMA, as measured by each scale, invariant?

RQ3. Is the correlation between general PERMA and workplace PERMA low enough to support distinctiveness?

2. Methods

2.1. Participants and Procedure

Participants were recruited via Amazon Mechanical Turk (MTurk). MTurk is an online crowdsourcing recruitment site for research. Individuals (aka MTurkers) create profiles based on their demographic, work, and other characteristics, and complete human intelligence tasks (HITs), which are commonly psychological studies that the individuals participate in. Researchers post studies on MTurk, and if MTurkers are eligible to participate (e.g., if their demographics match the inclusion criteria for a study), then they are able to enroll in the study and participate. If ineligible, they are not able to enroll. After participation, MTurkers' data are evaluated for quality by the study's research team. If acceptable, their HIT is approved; if unacceptable, their HIT is rejected. MTurkers' reputations (i.e., their proportion of approved HITs to total HITs) are tracked. English-speaking adults ($\geq$18 years) in the U.S. employed outside of MTurk were eligible to complete the surveys. Additionally, only high-reputation MTurkers (those with at least 95% approved tasks on MTurk) were allowed to participate [19]. Six hundred one participants were eligible, provided informed consent, and passed a majority of attention checks (3/5). See Table 1 for demographic characteristics.

Table 1. Demographic and descriptive statistics.

Variable	Mean (*SD*) or *n* (%)
Age (years)	35.39 (9.91)
Gender (female)	232 (38.60%)
Race/Ethnicity	
Black	80 (13.31%)
Native American	1 (0.17%)
White	411 (68.39%)
Asian	37 (6.16%)
Hispanic/Latino	40 (6.66%)
Multiracial	31 (5.16%)
Other/Missing Data	1 (0.17%)
Relationship Status	1 (7.14%)
Married	255 (42.43%)
Separated/Divorced	50 (8.32%)
Widowed	2 (0.33%)
Never Married	294 (48.92%)
Education	
High-School Diploma or GED	65 (10.87%)
Some College or Technical School	184 (30.77%)
Bachelor's Degree	262 (43.81%)
Some Graduate School	22 (3.68%)
Graduate or Professional Degree	65 (10.87%)
Income [a]	USD 62,000 (39,000)
PERMA-Profiler (range 0–10) [b]	
Positive Emotions	6.54 (2.40)
Engagement	6.75 (1.84)
Relationships	7.11 (2.42)
Meaning	6.87 (2.57)
Accomplishment	6.98 (2.00)
Workplace PERMA-Profiler (range 0–10) [b]	
Positive Emotions	6.22 (2.52)
Engagement	6.21 (2.36)
Relationships	7.07 (2.37)
Meaning	6.79 (2.58)
Accomplishment	7.35 (1.88)

Note. [a] Mean and standard deviation are rounded to the nearest thousand; [b] Means and standard deviations are computed using observed averages.

2.2. Measures

The PERMA-Profiler [1] and the Workplace PERMA-Profiler [2] were used to measure PERMA constructs in general and at work, respectively. In each scale, three items measured each construct on an 11-point scale (0: "Never" or "Not at All", to 10: "Always" or "Completely"): positive emotions (e.g., "How often do you feel positive?"; "At work, how often do you feel positive?"), engagement (e.g., "To what extent do you feel excited and interested in things?"; "To what extent do you feel excited and interested in your work?"), relationships (e.g., "How satisfied are you with your personal relationships?; "How satisfied are you with your professional relationships?"), meaning (e.g., "To what extent do you lead a purposeful and meaningful life?"; "To what extent is your work purposeful and meaningful?"), and accomplishment (e.g., "How often do you achieve the important goals you have set for yourself?"; "How often do you achieve the important work goals you have set for yourself?").

One notable divergence in item content appeared within a relationships item ("To what extent do you feel loved?" vs. "To what extent do you feel appreciated by your coworkers?"). McDonald's omega reliability coefficients for the PP were excellent for meaning ($\hat{\omega} = 0.94$) and relationships ($\hat{\omega} = 0.91$), good for positive emotions ($\hat{\omega} = 0.89$) and accomplishment ($\hat{\omega} = 0.87$), but poor for engagement ($\hat{\omega} = 0.66$). For the WPP, omega was excellent for meaning ($\hat{\omega} = 0.92$) and relationships ($\hat{\omega} = 0.91$) and good for positive emotions ($\hat{\omega} = 0.88$), accomplishment ($\hat{\omega} = 0.83$), and engagement ($\hat{\omega} = 0.82$). Omega for the overall flourishing factor was excellent ($\hat{\omega} = 0.0.96$) for both scales.

2.3. Statistical Analyses

RQ1. All models were evaluated using maximum likelihood estimation because the number of ordered categories exceeds seven (0–10)—which has previously been shown to equal that of least-squares estimation for ordinal data [20]—and all items were normally distributed. Models were identified using the marker indicator approach. First, model fit was evaluated for each scale independently. Model fit criteria included the Comparative Fit Index (CFI > 0.90), and the Standardized Root-Mean-Squared Residual (SRMR < 0.08), in accordance with Brown [21]. Next, I proceeded with single-sample invariance testing [21]. First, configural invariance was established by placing both scales into a single model and re-evaluating model fit using the same criteria. Metric invariance was assessed by constraining like-item loadings on the factors to be equivalent between each scale and evaluating the decrement in model fit between this and the configural model via the chi-squared test. Likewise, scalar invariance was assessed by constraining like-item intercepts to be equivalent between each scale and evaluating the model fit decrement between it and the metric invariance model via the chi-squared test. Partial invariance was evaluated in cases where full measurement invariance did not hold in either case [22]. Correlated residuals between like-items across scales were also included a priori in each model [21].

RQ2. Building on the scalar invariance model, latent factor variances were constrained to 1 in the WPP to be equivalent to the PP (which were set to 1 for model identification). A chi-squared difference test was computed to compare fit between this and the scalar invariant model. Next, covariances among latent factors within each scale were constrained to be equivalent across scales (e.g., the covariance between positive emotions and engagement constrained equal to the covariance between positive emotions and engagement in the WPP). Latent factor means were compared by evaluating the statistical significance ($p < 0.05$) of the mean difference between each like-construct across scales. Factor means of the PP were constrained to zero to identify the model, with those of the WPP freely estimated.

RQ3. The correlations between the PERMA like-factors across scales were evaluated from the covariance invariance model. If correlations were below 0.85, the latent factors were considered to be distinct across scales [21].

3. Results

See Table 2 for model fit and measurement invariance results. PERMA typically is a second-order factor, with positive emotions, engagement, relationships, meaning, and accomplishment as its first-order factors [3]. However, adding this second-order factor resulted in significant decrements in model fit with both the PP, $\Delta\chi^2(5, N = 601) = 72.04$, $p < 0.01$, and the WPP, $\Delta\chi^2(5, N = 601) = 112.20$, $p < 0.01$, indicating that a higher-order model would not be appropriate for analyses [21]. As such, only the first-order factors were retained. The fit of the five-factor PERMA model using the PP was excellent (CFI > 0.95; SRMR < 0.05). Although slightly worse, the fit of the model using the WPP was also good (CFI > 0.90; SRMR < 0.05). The model fit of the combined PERMA models with both scales was excellent (CFI > 0.95; SRMR < 0.05), supporting configural invariance (see Table 3 for the loadings and intercepts estimated in the configural model). Constraining the like-item loadings to be equivalent resulted in a significant decrement in model fit, $\Delta\chi^2(10, N = 601) = 115.33$, $p < 0.01$. Freeing three item loadings (one engagement item, one meaning item, and one accomplishment item) brought the fit reduction to an acceptable level, $\Delta\chi^2(7, N = 601) = 7.91$, $p = 0.34$, supporting partial metric invariance. As such, variances and covariances could reliably be compared across scales. Constraining all like-item intercepts to be equivalent also resulted in a significant decrement in model fit, $\Delta\chi^2(10, N = 601) = 88.72$, $p < 0.01$. Freeing five item intercepts (one positive emotion item, one engagement item, two relationship items, and one meaning item) brought the fit reduction to an acceptable level, $\Delta\chi^2(5, N = 601) = 10.54$, $p = 0.06$, supporting scalar invariance for all factors except for relationships. As such, the means for all but the relationships factor could be reliably compared.

Table 2. Model fit and invariance results.

Model	X^2	df	CFI	SRMR	$\Delta\chi 2$	Δdf	*p*-Value
PERMA-Profiler	425.90	80	0.959	0.033	--	--	--
Workplace PERMA-Profiler	495.43	80	0.948	0.041	--	--	--
Configural	1583.83	379	0.933	0.045	--	--	--
First-Order Metric	1695.86	389	0.927	0.056	112.03	10	<0.01
Second-Order Metric	1785.24	393	0.923	0.068	89.38	4	<0.01
Scalar	2044.46	408	0.909	0.076	259.22	15	<0.01
Partial Scalar	1208.52	357	0.953	0.040	10.54	5	0.06
Factor Variance	1260.16	362	0.950	0.049	51.64	5	<0.01
Partial Factor Variance	1215.73	360	0.952	0.041	7.21	3	0.06
Factor Covariance	1382.20	370	0.944	0.050	166.47	10	<0.01
Partial Factor Covariance	1215.76	361	0.953	0.041	0.03	1	0.86

Note. df is degrees of freedom. CFI is the Comparative Fit Index. SRMR is the Standardized Root-Mean-Squared Residual.

As indicated in Table 2, constraining all factor variances to 1 resulted in a significant decrement in model fit from the scalar invariance model, $\Delta\chi^2(5, N = 601) = 51.64$, $p < 0.01$. Freeing two factor variances from the WPP (engagement and meaning) brought the fit reduction to an acceptable level, $\Delta\chi^2(3, N = 601) = 7.21$, $p = 0.06$. Constraining all respective covariances to be equivalent across scales resulted in a significant decrement in model fit, $\Delta\chi^2(10, N = 601) = 166.47$, $p < 0.01$. Freeing all but one covariance constraint (positive emotions with relationships) brought the fit reduction to an acceptable level, $\Delta\chi^2(1, N = 601) = 0.03$, $p = 0.86$, indicating that the associations between PERMA factors were not equivalent across scales. Finally, all like-factor means except for relationships were compared across scales (because the relationships factor was not scalar-invariant). Workplace positive emotions ($M_{Diff} = -0.09$, $p = 0.01$), engagement ($M_{Diff} = -0.38$, $p < 0.01$), and meaning ($M_{Diff} = -0.10$, $p < 0.01$) were all significantly lower than in general. Workplace accomplishment, in contrast, was significantly higher than in general ($M_{Diff} = 0.19$, $p < 0.01$).

Table 3. Item characteristics across scales.

Factor Items [a]	PERMA-Profiler				Workplace PERMA-Profiler			
	U-Load	SE	Std Load	Intercept	U-Load	SE	Std Load	Intercept
Positive Emotions								
P1 [c]	2.25	0.09	0.82	6.01	2.42	0.10	0.81	5.43
P2	2.15	0.08	0.88	6.79	2.28	0.08	0.89	6.65
P3	2.37	0.09	0.85	6.82	2.42	0.10	0.84	6.56
Engagement								
E1 [c]	1.32	0.08	0.62	7.04	1.72	0.09	0.71	6.89
E2	2.22	0.09	0.88	6.96	2.79	0.10	0.93	6.18
E3 [b]	0.84	0.10	0.33	6.25	1.82	0.11	0.63	5.53
Relationships								
R1 [c]	1.91	0.09	0.75	6.97	2.03	0.09	0.82	7.16
R2 [c]	2.52	0.08	0.95	7.29	2.38	0.08	0.91	6.98
R3	2.48	0.09	0.91	7.09	2.39	0.08	0.91	7.04
Meaning								
M1	2.49	0.08	0.95	6.82	2.68	0.09	0.92	6.55
M2	2.36	0.09	0.87	6.91	2.55	0.09	0.90	6.76
M3 [b,c]	2.66	0.09	0.94	6.88	2.21	0.09	0.84	7.02
Accomplishment								
A1	2.13	0.08	0.89	6.51	2.12	0.08	0.89	6.84
A2	1.95	0.08	0.86	6.72	1.87	0.08	0.83	7.15
A3 [b]	1.54	0.07	0.73	7.72	1.11	0.08	0.56	8.03

Note. Coefficients in the table are taken from the configural invariance model. U-Load is the unstandardized loading. SE is the standard error of the unstandardized loading. Std Load is the standardizing loading. [a] Item numbers correspond to those published in [1,2]. [b] Items freed for partial metric invariance. [c] Items freed for partial scalar invariance.

See Table 4 for the correlations among latent factors. Although associations were high, particularly between accomplishment ($r = 0.81$) and positive emotions ($r = 0.81$), all like-construct correlations were under 0.85, indicating construct distinctiveness across scales.

Table 4. Latent variable correlations.

Variable	1	2	3	4	5	6	7	8	9
1. Positive Emotions	--								
2. W-Positive Emotions	0.81	--							
3. Engagement	0.96	0.83	--						
4. W-Engagement	0.70	0.91	0.79	--					
5. Relationships	0.86	0.73	0.83	0.61	--				
6. W-Relationships	0.67	0.86	0.67	0.76	0.61	--			
7. Meaning	0.87	0.75	0.90	0.71	0.75	0.74	--		
8. W-Meaning	0.69	0.90	0.75	0.96	0.61	0.78	0.74	--	
9. Accomplishment	0.91	0.79	0.93	0.73	0.76	0.78	0.93	0.74	--
10. W-Accomplishment	0.68	0.82	0.74	0.81	0.65	0.81	0.72	0.85	0.81

Note. "W-" indicates variables measured using the Workplace PERMA-Profiler. Other variables were measured with the PERMA-Profiler.

4. Discussion

When adapting scales for a single study, researchers may report the psychometric properties of their adaptation. However, without contextualizing with the parent scale's properties, it is unknown whether (1) the adaptation measures the intended construct in the same way as the parent scale (i.e., if the construct presents equivalently across contexts), and (2) whether the adaptation is statistically necessary (e.g., whether the inter-factor covariances and means are distinct). Previous meta-analyses have shown varying effects of interventions depending on whether a scale's adaptation or its parent was used [17,18], and there have been recent questions of the reliability of PERMA constructs measured with the

WPP [11,12]. A such, continuing evaluation of the PERMA scales' psychometric properties are needed to establish empirical support for their use in wider contexts like the workplace.

The results of the present study supported partial invariance across the PERMA-Profiler and the Workplace PERMA-Profiler, indicating that the flourishing constructs within PERMA were measured comparably across contexts. Not only does this finding support the consistent measurement structure of PERMA across contexts, but one could use these scales to directly compare the impacts of an intervention on flourishing both in general and at work. However, the relationships factor was not scalar-invariant, indicating that the means across these scales are not comparable. As such, using criteria derived from the PP to interpret means of the relationships factor from the WPP may not be appropriate. This lack of invariance could be due to the varied environmental setting (i.e., work), the variability of the individuals from whom the respondents are receiving support (i.e., others vs. coworkers), and the type of relational experience (i.e., love vs. appreciation). Each scale's structural covariances and latent means significantly differed, suggesting distinct factor means and inter-factor associations for PERMA constructs in each context. Additionally, correlations among like-constructs supported each scale's latent variable distinctiveness (e.g., workplace engagement and general engagement are not exactly the same).

4.1. Limitations and Future Directions

A number of limitations should be mentioned, as well as associated directions for future research. First, the study sample was recruited from MTurk was predominately male and educated and was analyzed cross-sectionally. There have been a number of studies on the quality, representativeness, and ethics of MTurk for survey research [19,23–26]. On average, MTurk samples have been found to be more representative of the general working population, and possibly more likely to provide valid data, than student samples [24] and professional panels [24,27], with some studies supporting normative sample alignment [28]. However, many papers cite problems with the reliability of responses, including inattentiveness, impossible answers, and an overrepresentation of white males [23,27,29,30]. Although I only included high-reputation workers, such efforts to increase the data quality may be futile as few researchers ever reject HITs [31]. As such, our results may be biased by poor data quality and may not be generalizable to all worker populations, especially those outside of the United States. Replications of this study are recommended using different samples, preferably those recruited directly from workplaces, and with a larger cross-national and female proportion.

Another major limitation is the lack of associative variables for convergent validity comparisons. Previous research has shown stronger associations between workplace interventions and workplace contextual measures than general measures (e.g., workplace mental health stigma vs. general mental health stigma; [17]). As such, a more robust evaluation of the PP's and WPP's distinct utility would have involved the structural invariance of PERMA constructs with practical workplace measures (e.g., job performance, satisfaction, burnout [9,32,33]). Future research should evaluate the practical utility of the WPP over the PP in workplace contexts through additional structural invariance assessments. Such results would provide leaders with more evidence for the utilization of one scale over the other for workplace-specific emotional wellbeing pulse-taking.

A final future research direction would be to continue evaluating the psychometric properties of both scales. Recent research has suggested poorer reliability than is acceptable with the PP and WPP [11,12]. In the present study, I too found unacceptably low internal consistency for the engagement subscale of the PP, and the relationships subscale had non-invariant measurement properties, suggesting that (1) our results may not hold across replications, and (2) interpretations of relationships and the higher-order PERMA factor may not be comparable across general and workplace contexts, and consequently the use of either scale may not be recommendable to evaluate relationship-related health at work. I recommend adjustments to the engagement items be made to improve reliability, and adjustments to the relationship items be made (perhaps solely for the WPP) so that

these constructs may be more representative of the same factor. Future work should be carried out to develop these adjustments and test the psychometric differences between the subsequent adaptation and the original versions.

4.2. Conclusions

When measuring workplace wellbeing, it is important that the scale(s) used are placed in a workplace context. However, if an adaptation to a general scale is made in order to do so, psychometric evaluations of the new scale are necessary to ascertain measurement consistency with and structural distinction from its parent scale. Without such tests, it is unknown whether the results are interpretable in the same way, and such practical implications may be questionable. Our results suggest that the measurement of PERMA is relatively stable across scales, and the constructs themselves are distinct, supporting the utility of both the PERMA-Profiler and the Workplace PERMA-Profiler. However, I express caution when measuring the relationships factor in the workplace (and, consequently, if the higher-order PERMA factor is used), due to its non-invariance between general and workplace contexts. The practical implications of this include support of the WPP for leaders to use as an employee emotional health assessment.

References

1. Butler, J.; Kern, M.L. The PERMA-Profiler: A brief multidimensional measure of flourishing. *Int. J. Wellbeing* **2016**, *6*, 1–48. [CrossRef]
2. Kern, M.L. The Workplace PERMA Profiler. 2013. Available online: https://www.peggykern.org/uploads/5/6/6/7/56678211/workplace_perma_profiler_102014.pdf (accessed on 3 March 2024).
3. Seligman, M.E.P. *Flourish: A Visionary New Understanding of Happiness and Well-Being*; Simon & Schuster: New York, NY, USA, 2011.
4. Dagenais-Desmarais, V.; Savoie, A. What is psychological well-being, really? A grassroots approach from the organizational sciences. *J. Happiness Stud.* **2012**, *13*, 659–684. [CrossRef]
5. Luthans, F.L.; Avolio, B.J.; Avey, J.A. *Psychological Capital Questionnaire (PsyCap)*; [Database Record]; APA PsycTests: 2007. Available online: https://doi.org/10.1037/t06483-000 (accessed on 3 March 2024).
6. Lorenz, T.; Beer, C.; Pütz, J.; Heinlitz, K. Measuring psychological capital: Construction and validation of the Compound PsyCap Scale (CPC-12). *PLoS ONE* **2016**, *11*, e0152892. [CrossRef]
7. Culbertson, S.S.; Fullagar, C.J.; Mills, M.J. Feeling good and doing great: The relationship between psychological capital and well-being. *J. Occup. Health Psychol.* **2010**, *15*, 421–433. [CrossRef] [PubMed]
8. Ryff, C.D. Happiness is everything or is it? Explorations on the meaning of psychological wellbeing. *J. Personal. Soc. Psychol.* **1989**, *57*, 1069–1081. [CrossRef]
9. Watanabe, K.; Kawakami, N.; Adachi, H.; Matsumoto, K.; Imamura, K.; Matsumoto, K.; Yamagami, F.; Fusejima, A.; Muraoka, T.; Kagami, T.; et al. The Japanese Workplace PERMA-Profiler: A validation study among Japanese workers. *J. Occup. Health* **2018**, *60*, 383–393. [CrossRef] [PubMed]
10. Choi, S.P.; Suh, C.; Yang, J.W.; Ye, B.J.; Lee, C.K.; Son, B.C.; Choi, M. Korean translation and validation of the Workplace Positive emotion, Engagement, Relationships, Meaning, and Accomplishment (PERMA)-Profiler. *Ann. Occup. Environ. Med.* **2019**, *31*, e17. [CrossRef] [PubMed]

11. Jimenez, W.P.; Hu, X.; Garden, R.; Xie, X. Toward a more PERMA(nent) conceptualization of worker well-being? A cross-cultural study of the Workplace PERMA-Profiler. *J. Pers. Pscyhol.* **2021**, *21*, 94–100. [CrossRef]
12. Jimenez, W.P.; Hu, X.; Garden, R.; Zeytonli, A. The potential and peculiarities of PERMA: A meta-analysis of two well-being measures with working samples. *J. Pers. Psychol.* **2024**, *23*, 49–57. [CrossRef]
13. Holder, M.D.; Klassen, A. Temperament and happiness in children. *J. Happiness Stud.* **2010**, *11*, 419–439. [CrossRef]
14. Lyubomirsky, S.; Lepper, H.S. A measure of subjective happiness: Preliminary reliability and construct validation. *Soc. Indic. Res.* **1999**, *46*, 137–155. [CrossRef]
15. Tkach, C.; Lyubomirsky, S. How do people pursue happiness?: Relating personality, happiness-increasing strategies, and well-being. *J. Happiness Stud.* **2006**, *7*, 183–225. [CrossRef]
16. Layous, K.; Nelson, S.K.; Oberle, E.; Schonert-Reichl, K.A.; Lyubomirsky, S. Kindness counts: Prompting prosocial behavior in preadolescents boosts peer acceptance and well-being. *PLoS ONE* **2012**, *7*, e51380. [CrossRef] [PubMed]
17. Tóth, M.D.; Ihionvien, S.; Leduc, C.; Aust, B.; Amann, B.L.; Cresswell-Smith, J.; Reich, H.; Cully, G.; Sanches, S.; Fanaj, N.; et al. Evidence for the effectiveness of interventions to reduce mental health related stigma in the workplace: A systematic review. *BMJ Open* **2023**, *13*, e067126. [CrossRef] [PubMed]
18. Di Sante, J.; Akeson, B.; Gossack, A.; Knäuper, B. Efficacy of ACT-based treatments for dysregulated eating behaviours: A systematic review and meta-analysis. *Appetite* **2022**, *171*, 105929. [CrossRef] [PubMed]
19. Peer, E.; Vosgerau, J.; Acquisti, A. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behav. Res. Methods* **2014**, *46*, 1023–1031. [CrossRef] [PubMed]
20. Rhemtulla, M.; Brosseau-Liard, P.E.; Savalei, V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol. Methods* **2012**, *17*, 354–373. [CrossRef] [PubMed]
21. Brown, T.A. *Confirmatory Factor Analysis for Applied Research*, 2nd ed.; Guilford Press: New York, NY, USA, 2015.
22. Vandenberg, R.J.; Lance, C.E. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations. *Organ. Res. Methods* **2000**, *3*, 4–70. [CrossRef]
23. Chmielewski, M.; Kucker, S.C. An MTurk crisis? Shifts in data quality and the impact on study results. *Soc. Psychol. Personal. Sci.* **2019**, *11*, 464–473. [CrossRef]
24. Kees, J.; Berry, C.; Burton, S.; Sheehan, K. An analysis of data quality: Professional panels, student subject pools, and Amazon's Mechanical Turk. *J. Advert.* **2017**, *46*, 141–155. [CrossRef]
25. Moss, A.J.; Rosenzweig, C.; Robinson, J.; Jaffe, S.N.; Litman, L. Is it ethical to use Mechanical Turk for behavioral research? Relevant data from a representative survey of MTurk participants and wages. *Behav. Res. Methods* **2023**, *55*, 4048–4067. [CrossRef] [PubMed]
26. Woo, S.F.; Keith, M.; Thornton, M.A. Amazon Mechanical Turk for industrial and organizational psychology: Advantages, challenges, and practical recommendations. *Ind. Organ. Psychol.* **2015**, *8*, 171–179. [CrossRef]
27. Smith, S.M.; Roster, C.A.; Golden, L.L.; Albaum, G.S. A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples. *J. Bus. Res.* **2016**, *69*, 3139–3148. [CrossRef]
28. Coppock, A. Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. *Political Sci. Res. Methods* **2019**, *7*, 613–628. [CrossRef]
29. Nadler, J.; Baumgartner, S.; Washington, M. MTurk for working samples: Evaluation of data quality 2014–2020. *N. Am. J. Psychol.* **2021**, *23*, 741–752.
30. Rouse, S.V. A reliability analysis of Mechanical Turk data. *Comput. Hum. Behav.* **2015**, *43*, 304–307. [CrossRef]
31. Hauser, D.J.; Moss, A.J.; Rosenzweig, C.; Jaffe, S.N.; Robinson, J.; Litman, L. Evaluating CloudResearch's Approved Group as a solution for problematic data quality on MTurk. *Behav. Res. Methods* **2023**, *55*, 3953–3964. [CrossRef] [PubMed]
32. Bazargan-Hejazi, S.; Shirazi, A.; Wang, A.; Shlobin, N.A.; Karunungan, K.; Shulman, J.; Marzio, R.; Ebrahim, G.; Shay, W.; Slavin, S. Contribution of a positive psychology-based conceptual framework in reducing physician burnout and improving well-being: A systematic review. *BMC Med. Educ.* **2021**, *21*, 593. [CrossRef]
33. Kern, M.L.; Waters, L.; Adler, A.; White, M. Assessing employee wellbeing in schools using a multifaceted approach: Associations with physical health, life satisfaction, and professional thriving. *Psychology* **2014**, *5*, 500–513. [CrossRef]