

From multiverse analysis to multiverse operationalisations: 262,143 ways of measuring well-being

Paul H. P. Hanel & Natalia Zarzeczna

To cite this article: Paul H. P. Hanel & Natalia Zarzeczna (2022): From multiverse analysis to multiverse operationalisations: 262,143 ways of measuring well-being, Religion, Brain & Behavior, DOI: [10.1080/2153599X.2022.2070259](https://doi.org/10.1080/2153599X.2022.2070259)

To link to this article: <https://doi.org/10.1080/2153599X.2022.2070259>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 06 Jul 2022.



[Submit your article to this journal](#)



Article views: 50




[View related articles](#)



[View Crossmark data](#)

From multiverse analysis to multiverse operationalisations: 262,143 ways of measuring well-being

Paul H. P. Hanel ^a and Natalia Zarzeczna^b

^aDepartment of Psychology, University of Essex, Colchester, UK; ^bDepartment of Psychology, University of Amsterdam, Amsterdam, Netherlands

Researchers are increasingly noticing the value of multiverse analysis. While such analysis is important, they should not neglect the impact of multiverse operationalisations on the outcome—that is, multiple ways of operationalizing or measuring constructs such as well-being or religiosity. The “Many Analysts Religion Projects” (2022) found that different operationalisations of well-being impacted its association with religiosity. Here, we expand on this by comparing how 262,143 ways of measuring well-being are associated with religiosity and 511 ways of measuring religiosity are associated with well-being. We found that especially short 1–2 item measures of well-being that aimed to capture it broadly resulted in a spread of associations ranging roughly between $r = 0-.25$. This suggests that depending on the operationalization of well-being, researchers might under- or overestimate the strengths of its association with religiosity. Thus, we recommend that different measures of the same construct are more routinely included in studies to gauge the robustness of the finding.


In recent years, researchers found that the analytic method can impact the results (Botvinik-Nezer et al., 2020; Schweinsberg et al., 2021; The MARP team, 2022). For example, Silberzahn et al. (2018) found that the finding showing a negative bias against American-footballers with darker skin tones depended on the statistical analysis. However, it has been argued that it is, in fact, the interpretation of a research question that leads to inconsistent research outcomes (Auspurg & Brüderl, 2021).

Indeed, the MARP team (2022) found that analytic choice *and* operationalization mattered. For example, how well-being was operationalized impacted its association with religiosity: Psychological well-being was the most strongly and physical well-being weakly associated with religiosity even though internal consistencies were satisfactory for all scales. However, researchers involved in the project used only a small proportion of the overall amount of 262,143 possible operationalisations.¹

To the best of our knowledge, only one article has systematically investigated whether the number of items in a scale impacts the criterion validity. Bakker and Lelker (2018) found that when need for cognition is measured with a longer scale, its correlation with relying on policy information is nearly twice as large as when it is assessed with a shorter scale.

In the present study, we test whether the operationalization of well-being had an impact on its association with religiosity. In other words, we kept the analysis (Pearson’s correlation) constant and only varied the operationalization. Additionally, we tested to what extent operationalization of religiosity impacted the association with well-being. While it has already been established that there are many ways to measure well-being (Cooke et al., 2016), it is important to systematically

CONTACT Paul H. P. Hanel  p.hanel@essex.ac.uk

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/2153599X.2022.2070259>

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

test how large the impact is. The R-code to reproduce our analysis, data, and figures can be found in the Supplementary Materials.

Method

We used the data provided by the initiators of the MARP team (2022), which consists of 10,535 participants from 24 countries. We excluded 340 participants prior to any analysis for failing an attention check ($M_{\text{age}} = 33.76$, $SD = 13.82$). Including them did not change the pattern of results. Since performing analyses with all participants required too much processing power, we only selected participants from Germany, as it was the country with most participants ($n = 1,287$). We replicated the analyses with data from the largest Asian country included in the dataset (Japan, $n = 424$) and American country (Brazil, $n = 402$).

Participants responded to a range of measures. Of these, we only use well-being and religiosity. Well-being was measured with an 18-item scale ($\alpha = .90$) that consisted of three dimensions: physical well-being (e.g., “Do you have enough energy for everyday life?”; 7-items, $\alpha = .77$), psychological well-being (e.g., “How satisfied are you with yourself?”; 6-items, $\alpha = .82$), and social well-being (“How satisfied are you with your personal relationships?”; 3-items, $\alpha = .70$). Additionally, there were two items to assess general well-being (e.g., “How would you rate your quality of life?”). Religiosity was measured with 9-items such as “How often do you pray?” and “To what extent do you consider yourself to be spiritual?” ($\alpha = .93$).

Results

We operationalized well-being by using all possible combinations of the 18-items. For example, we computed the mean of all 153 possible combinations of 2-items and correlated those 153 2-item well-being scales with the overall religiosity score. Non-surprisingly, combining more items were associated with a smaller spread in correlation coefficients (Figure 1). For example, whereas the

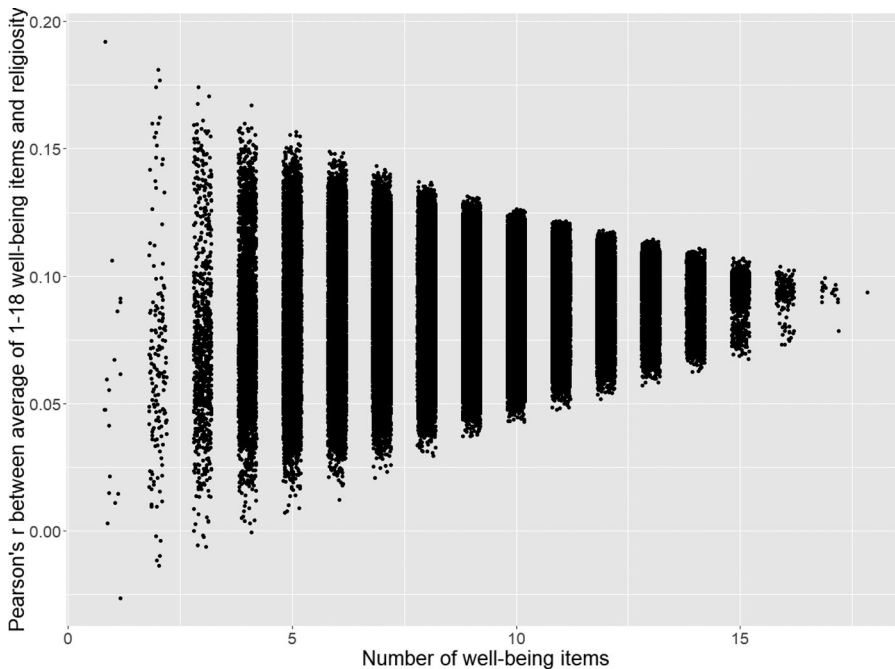


Figure 1. Correlations between 262,143 operationalisations of well-being with religiosity.

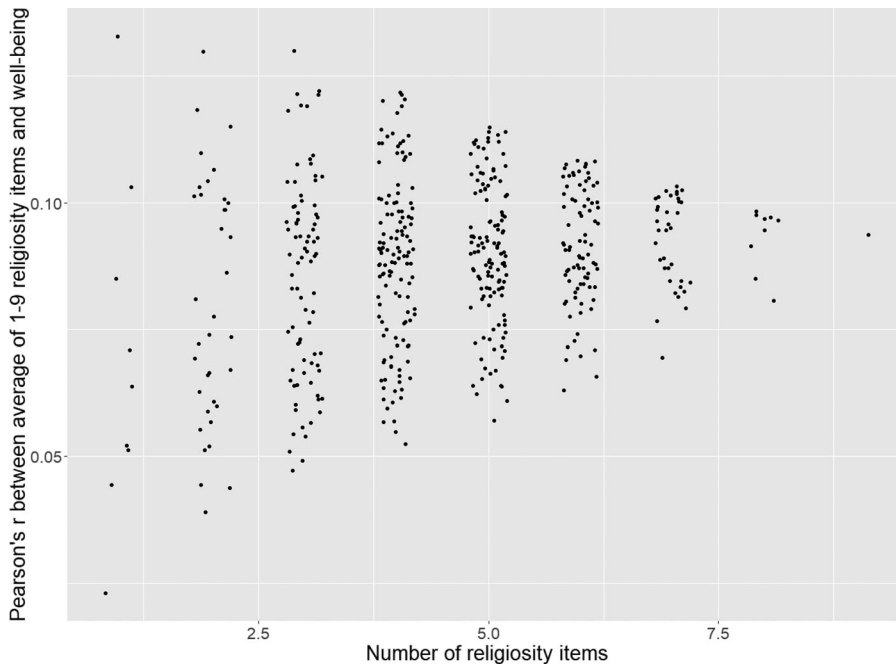


Figure 2. Correlations between 511 operationalisations of religiosity with well-being.

153 correlations of all two-item measures of well-being with religiosity ranged from $r = -.01 - .18$, the 18 correlations of the 17-item measure ranged only from $r = .08-.10$. The spread was reduced when we focused on physical, psychological, and social well-being separately (Figures S1-S3). The spread was wider in samples of 424 Japanese and 402 Brazilian respondents (Figures S4-11).

We then tested whether 511 operationalisations of religiosity impacted the association with the 18-item well-being index. The spread was smaller, suggesting that the religiosity items are more homogenous than the well-being items (see Figure 2, for replications from Japan and Brazil, see Figures S12-S13).

Discussion

In the present manuscript, we investigated how different operationalisations of well-being and religiosity impact the association strength between these two constructs. Researchers argued that measuring broad constructs with short scales is acceptable even if it comes with lower internal consistencies. This is because such scales take less time to complete while still capturing the depth of the construct (Gosling, *n.d.*; Graham et al., 2011). Our results point to potential challenges of this approach. Using data collected by the MARP team, we computed correlations between religiosity and well-being using every possible way of operationalizing the constructs (e.g., we computed a scale of well-being by averaging every 2-items used to measure well-being. Then, we correlated the 2-item mean scores with the overall religiosity score, which was operationalized by averaging all items that measured religiosity). We found that measuring well-being as a broad construct (with only a few items) can lead to a wider spread of the findings. When measuring specific facets of well-being, such as psychological or social well-being, this is less of an issue.

It is worth noting that single items can have a strong impact. Figure 1 shows that in the German sample, one well-being item is clearly more strongly correlated with religiosity than the other items. This item is part of the psychological well-being facet: “To what extent do you feel your life to be meaningful?” It correlates with religiosity with $r = .19$ in the German sample ($r = .24$ in the Japanese

sample and $r = .37$ in the Brazilian sample, cf. Figures S4 and S8). Non-surprisingly, excluding this item has relatively the strongest impact on the overall association of the remaining 17 well-being items with religiosity in all three samples. Also, excluding the meaningfulness-item from the overall well-being or psychological well-being scale would have resulted in a lower Cronbach's α s, suggesting that the internal consistency might not be an informative indicator of criterion validity.

Of course, it is also a theoretical question whether certain items measure well-being. For example, it has been argued that meaningfulness is theoretically distinct from other aspects of (psychological) well-being such as satisfaction with life, happiness, and positive affect (Crego et al., 2020). Nevertheless, the meaningfulness item was the item that was most often included as dependent variable for both research questions by all MARP-teams (2022). While this seems reasonable, it should be supplemented by an item-by-item analysis to get a better understanding which well-being facets are stronger correlated with religiosity.

Our finding that meaningfulness is more strongly associated with religiosity than other aspects of well-being is broadly in line with meta-analytic evidence (Hackney & Sanders, 2003). However, unlike meta-analyses that mostly focus on scale—or construct-level effects, our approach allows to assess the impact of each item. Of course, an item-by-item analysis is less reliable than using validated scales and therefore replication is even more important. In samples from three countries, we found that meaningfulness was most strongly associated with religiosity, thereby increasing our trust in the outcome.

Different operationalisations of religiosity had a less strong impact on the association with well-being. This suggests that religiosity might be a more homogeneous construct than well-being.

Together, our results expand the findings of the MARP team (2022) by quantifying the extent to which differences in operationalization impact the outcome. We recommend that researchers only use shorter scales if they have been well validated and otherwise use longer and more comprehensive scales. Researchers should also consider using multiple scales for each of the key constructs to test whether their effects are measure-independent.

Note

1. The survey included 18 well-being items. There are $C(18, 1) = 18$ possibilities to select 1 item, $C(18, 2) = 153$ to select two items, ..., $C(18, 9) = 48,620$ possibilities to select 9 items etc. resulting in 262,143 operationalisations in total.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Paul H. P. Hanel  <http://orcid.org/0000-0002-3225-1395>

References

- Auspurg, K., & Brüderl, J. (2021). Has the credibility of the social sciences been credibly destroyed? Reanalyzing the “Many Analysts, One data Set” project. *Socius: Sociological Research for a Dynamic World*, 7. <https://doi.org/10.1177/23780231211024421>
- Bakker, B. N., & Leikes, Y. (2018). Selling ourselves short? How abbreviated measures of Personality change the Way We think about Personality and politics. *The Journal of Politics*, 80(4), 1311–1325. <https://doi.org/10.1086/698928>
- Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., ... Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 582(7810), 84–88. <https://doi.org/10.1038/s41586-020-2314-9>

- Cooke, P. J., Melchert, T. P., & Connor, K. (2016). Measuring well-being: A review of instruments. *The Counseling Psychologist, 44*(5), 730–757. <https://doi.org/10.1177/0011000016633507>
- Crego, A., Yela, J. R., Gómez-Martínez, M.Á., & Karim, A. A. (2020). The contribution of meaningfulness and mindfulness to psychological well-being and Mental Health: A structural equation model. *Journal of Happiness Studies, 21*(8), 2827–2850. <https://doi.org/10.1007/s10902-019-00201-y>
- Gosling, S. D. (n.d.). A note on alpha reliability and factor structure in the TIPI. <http://gosling.psy.utexas.edu/scales-weve-developed/ten-item-personality-measure-tipi/a-note-on-alpha-reliability-and-factor-structure-in-the-tipi/>.
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101*(2), 366–385. <https://doi.org/10.1037/a0021847>
- Hackney, C. H., & Sanders, G. S. (2003). Religiosity and mental health: A meta-analysis of recent studies. *Journal for the Scientific Study of Religion, 42*(1), 43–55. <https://doi.org/10.1111/1468-5906.t01-1-00160>
- Schweinsberg, M., Feldman, M., Staub, N., van den Akker, O. R., van Aert, R. C. M., van Assen, M. A. L. M., Liu, Y., Althoff, T., Heer, J., Kale, A., Mohamed, Z., Amireh, H., Venkatesh Prasad, V., Bernstein, A., Robinson, E., Snellman, K., Amy Sommer, S., Otner, S. M. G., Robinson, D., ... Luis Uhlmann, E. (2021). Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. *Organizational Behavior and Human Decision Processes, 165*, 228–249. <https://doi.org/10.1016/j.obhdp.2021.02.003>
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š, Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., ... Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science, 1*(3), 337–356. <https://doi.org/10.1177/2515245917747646>
- The MARP team. (2022). A many analysts approach to the relation between religiosity and well-being. *Manuscript in preparation*. <https://osf.io/qbdce/>.