

Evidence of Validity Does not Rule out Systematic Bias: A Commentary on Nomological Noise and Cross-Cultural Invariance

Sociological Methods & Research

2023, Vol. 52(3) 1420–1437

© The Author(s) 2022

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/00491241221091756

journals.sagepub.com/home/smr

Ronald Fischer ^{1,2},
Johannes Alfons Karl ¹,
Johnny R. J. Fontaine ³, and
Ype H. Poortinga⁴

Abstract

We comment on the argument by Welzel, Brunkert, Kruse and Inglehart (2021) that theoretically expected associations in nomological networks should take priority over invariance tests in cross-national research. We agree that narrow application of individual tools, such as multi-group confirmatory factor analysis with data that violates the assumptions of these techniques, can be misleading. However, findings that fit expectations of nomological networks may not be free of bias. We present supporting evidence of systematic bias affecting nomological network relationships from a) previous research on intelligence and response styles, b) two simulation studies, and c) data on the choice index from the World Value Survey

¹Victoria University of Wellington, New Zealand

²Institute D'Or for Research and Teaching, Rio de Janeiro, Brazil

³Department of Work, Organization and Society, Ghent University, Belgium

⁴Tilburg University, The Netherlands

Corresponding Author:

Ronald Fischer, School of Psychology, Victoria University of Wellington, New Zealand.

Email: ronald.fischer@vuw.ac.nz

(WVS). Our main point is that nomological network analysis by itself is insufficient to rule out systematic bias in data. We point out how a thoughtful exploration of sources of biases in cross-national data can contribute to stronger theory development.

Keywords

Invariance, nomological networks, systematic bias, values, cross-cultural differences, multilevel models, simulation, choice index

The recent paper by Welzel and colleagues (Welzel et al., 2021) argues that multi-group confirmatory factor analysis (MGCFA) is an inappropriate method for testing non-invariance, or “cross-cultural bias”, because of problems with this method when it is used for assessing constructs that show large mean differences and very skewed distributions at the extremes in cross-national data sets. We agree that MGCFA can impose too strict conditions. Extending from these observations, Welzel et al. argue that non-invariance is inconsequential for cross-cultural analyses as long as the nomological network of theoretically expected relationships with the target variable is meaningful. We believe this claim is not justified in cross-national research including research on social values. To demonstrate our point, we present evidence for systematic bias in nomological network analyses based on past research, simulations, and real data. We emphasize that invariance testing, independent of the chosen method, should be an integral part of cross-national research, both to reduce the risk of invalid interpretations and to advance meaningful theoretical developments. The identification of invariance should not be seen as a purely “methodological gatekeeper”, but also provides the basis for a richer exploration of how theoretically predicted relationships may operate and thereby offer an opportunity to enrich our understanding of how cultural dynamics operate at both individual and societal levels.

We focus on the two major points within the target article and restate these here for the sake of common understanding:

- (a) Welzel et al. criticise the use of MGCFA as a methodological gatekeeper for non-invariance in the case of constructs on which extreme differences can be observed between countries and there

are floor and ceiling effects in the items that are used to operationalize these constructs.

- (b) They claim that the observation of a theoretically expected nomological network at country level takes priority over psychometric evidence of bias in the assessment procedures within each of the countries.

Frontiers and opportunities for invariance testing

We agree with Welzel et al. that MGCFA can lead to a misleading evaluation of equivalence in a cross-country data set when assumptions about distributions of variables are grossly violated, as is the case with score distributions that are extremely skewed. It is important to note that MGCFA is only one tool within a larger methodological tool kit that can be applied to test invariance or equivalence of measures (Boer et al., 2018; Fischer & Karl, 2019; Matsumoto & van de Vijver, 2011). There are further problems with the broader latent variable approach, that include MGCFA models. Other authors have addressed issues on top of those noted by Welzel et al. (e.g., Barrett, 2007; Borsboom et al., 2003; Chen, 2007; Dalege et al., 2016; Kang et al., 2016; Marsh et al., 2004). Limitations of latent variable approaches in large-scale analyses with multi-item indicators are well documented and research is progressing to address issues that may help contextualize any invariance identified. Two areas of particular relevance in this regard we believe merit further attention. The first are attempts to identify effect size measures of lack of invariance which specify the magnitude of invariance parameters, instead of relying solely on discrete bias/no-bias decisions based on significance levels and traditional fit criteria (Nye & Drasgow, 2011). The second area consists of theory-driven multilevel approaches that identify sources of misfit (Davidov et al., 2012; Fischer et al., 2011; Fontaine et al., 2008; van de Vijver & Poortinga, 2002). The focus of these models is to understand and explain the sources of misfit, which can provide novel insights into both measurement and culture theory (see below). Furthermore, it is important to reiterate that MGCFA is just one specific tool within a larger arsenal of invariance tests (see Boer et al., 2018; Fischer & Karl, 2019) and that invariance testing provides opportunities for more nuanced insights into social dynamics. It is theoretically meaningful to explore the relative size and sources of invariance. We strongly recommend more attention to using diverse invariance tools, building on the respective strengths of each method.

Nomological networks and nomological noise

Our focus in the current response is on the recommendation for researchers ‘to give up comparing group-wise dimensional statistics in cases in which group means are so disparate that a substantial number of them reaches into the floor and ceiling section of a closed-ended scale’ (pp. 23). Instead, nomological network analyses are recommended. We strongly believe that this recommendation is ill-advised and detrimental to the larger field of cross-national (values) research. Nomological network analyses are prone to suffer from suppression effects¹; these are common and often not recognized or acknowledged (MartinezGutierrez & Cribbie, 2021). In addition, we argue that this recommendation overlooks the possibility of what has been coined in the literature as “nomological noise” (Messick, 1981, 1989). Nomological noise refers to the phenomenon that (a part of) the relationships in the nomological network are to be attributed to other factors than the intended ones (e.g., the presence of an acquiescence [Yes-saying tendency] or social desirability factor that artificially inflates associations between variables within the network). These method effects can also affect the nomological network and need to be identified. We present three different lines of evidence to show why nomological noise is important to rule out when examining nomological network relationships at the societal level.

First, we review evidence from intelligence and response styles research demonstrating that nomological networks at the country level can show interpretable correlations which can be explained by clear method artefacts at the individual level. Second, we report two simulation studies (building on the work by (Alwin, 2007), cited by Welzel et al.), using a classic bi-factor model (Welkenhuysen-Gybels et al., 2003) at the individual level, without an explicit model for the nation-level mean structure. These simulations demonstrate that systematic method variance only at the individual level can bias correlations at the individual level *and* can have follow-on effects at the nation-level. Third, we use the choice index from the WVS to show how systematic variance at the nation level affects both internal and external validity coefficients at individual and nation-level.

Previous research on bias in scores on psychometric tests and questionnaires

When examining the consequences of ignoring cross-cultural invariance testing, the most notorious and well-documented cases of systematic method biases come from the history of intelligence testing. The claim that

population differences were valid derived from the strong nomological network surrounding IQ batteries at the individual level. Various kinds of ability tests showed a consistent correlational structure, supporting a hierarchical model with a large number of task-specific factors at the lowest level, a limited number of broad ability factors at the second level, and a general intelligence factor (called “g”) at the top of the conceptual hierarchy (Carroll, 1993). Within a population this model works well; individual differences in IQ predict successfully socially important criteria, such as academic success and professional achievement (Schmidt & Hunter, 1998). When applied across populations the model *seemingly* works well too, showing similar predictive evidence. The inescapable failure of the model across populations that differ in background is demonstrated by method effects that vary both within and across populations, i.e., test-bound sources of variance that are systematic, but are not pertinent to the target construct of a measure. For example, Malda et al. (2010) developed five cognitive tests for the Tswana population and for the Afrikaner population in South Africa, using item content with which children were familiar. Children generally performed better on the test version designed for their own group than on the other version of the test. Effects of familiarity with task content have been demonstrated by Cole and co-workers across a wide range of populations and for a range of cognitive tasks (Scribner & Cole, 1981) and have been long documented in related areas of perceptual skills (Serpell, 1979). Lack of familiarity is a method factor that can affect IQ scores of individual test takers within nations.

Distortions of score differences are not limited to tests where correct answers and incorrect answers are clearly distinguished. Lack of familiarity with response options used in frequently used Likert scales can have a dramatic effect on the endorsement of the response options (Weijters et al., 2013). The most widely studied sources of systematic bias in value scales and other survey scales are social desirability and response styles (Harzing, 2006; He et al., 2015; Smith et al., 2016). Social desirability responding varies both within and across countries and it is possible to use a social desirability scale as a standard against which social desirability responding in target variables can be assessed (He et al., 2015). Similarly, differences in the frequency of selected response scale categories provide evidence of extreme and acquiescent responding (Harzing, 2006). Such differences in acquiescence when responding to surveys can misrepresent actual differences in overt behavior. van Herk et al. (2004) demonstrated this with data from marketing surveys in six European countries in which acquiescence scores could be linked to national consumer statistics and self-reported actual

behavior. In this study, country differences in survey data did not match actual volumes of sales. Rather than being unsystematic, social desirability and response styles such as acquiescence tend to differ systematically across countries and even have been interpreted as a characteristic of collectivist as opposed to individualist societies (Smith, 2004; Smith et al., 2016). In summary, there are plausible method factors that contribute to systematic correlations between weakly or unrelated tests at societal level and may enhance or obscure differences between populations.

The presence of bias works in two ways: it challenges interpretations of observed differences in scores between populations as if they were equivalent, but equally precludes the conclusion that there are no differences between populations. This is the position of Van de Vijver and colleagues, who long before MGCFA became popular, distinguished between levels of equivalence (Van De Vijver & Poortinga, 1982; van de Vijver & Leung, 1997) and between individual-level and population-level interpretation (Van De Vijver et al., 2008). They did not ‘idealize’ measures that show little variation between groups, as suggested by Welzel et al. (2021, p. 15). In fact, van de Vijver (2011) emphasized that bias (i.e., lack of equivalence) and valid differences should not be seen as antithetical but as complementary sources of observed cross-cultural differences in measured scores. Van de Vijver mentioned a study by Fontaine et al. (2008) that demonstrated a good average representation of the Schwartz value theory across samples, but at the same time showed a pattern of systematic deviations from this structure. For more affluent countries there was a larger contrast between protection and growth values compared to less affluent countries. Thus, the size of the structural bias at the individual within country-level was systematically related to country-level characteristics.

Summarizing our brief literature review, there is clear empirical evidence across very different types of measurement that method factors can vary systematically across individuals as well as across groups and can lead to nomological noise at the group-level. This diverse literature thus demonstrates that observing theoretically expected relations between variables in a nomological network does not exclude the possibility of systematic method artefacts.

Systematic method variance can bias nomological network correlations: A simulation

To examine whether systematic bias in observed indicators at the individual level can spill over to the population (or nation-) level, we set up a simulation

study mimicking a common case in self-report research, namely individual differences in acquiescent responding (see Supplement A for technical details). The simulation used a classic individual-level bi-factor design to model systematic method variance independent of the substantive factors (Rodriguez et al., 2016; Welkenhuysen-Gybels et al., 2003). In a bi-factor model, individual responses are thought to be related to substantive factors as in standard psychometric analyses as well as an additional factor which affects all items equally, but is unrelated to the substantive factors. We used this set-up to examine the impact of an individual level method effect uncorrelated with the substantive variables in a multi-level environment (see Figure 1).

We conducted two different simulations to examine the effects of different strengths of this systematic bias factor at the individual level on the observed correlations at both the individual and the societal level (see Sokolov, 2018 for a simulation demonstrating method effects at the nation-level). In the first simulation, we compared four different conditions. We used a condition in which the bi-factor had no influence on the individual items and three conditions with increasing bias (from 0.10 to 0.30 to 0.50). To put this in context, the factor loadings on the substantive factor were set via a reference loading of 1.00; the effect of the bi-factor on the items was therefore substantively lower than the two substantive factors.

We created 5,000 datasets for each of the four conditions (including 500 observations in each dataset), saved the item means from each simulated data set, averaged the item means per factor and then correlated the means at the nation level. Comparing the four conditions, we found a strong effect of the systematic bias variable on the averaged correlations per condition: $F(3,396) = 14.62$, $p < .001$, $\eta^2 = .099$ (see Figure 2). Increasing bias at individual level was associated with inflated correlations between the item averages at the societal level. It is important to emphasize that we did not model a mean-level structure, we only systematically varied the influence of an individual-level variable and observed the follow-on effects at the nation level. As the strength of the bi-factor on the items increased, the nation-level correlations became stronger and deviated from the simulated 'true' correlation. We also set up a more complex simulation which examined a wider range of bias options, which allowed us to examine both the individual-level and nation-level correlation patterns depending on the strength of the bias factor at the individual level (see Supplement A for more information). The main take-away point of both versions of this simple simulation is that examining the internal validities at the individual level is important and researchers should consider and empirically test for the presence of systematic bias at this level when interpreting population differences.

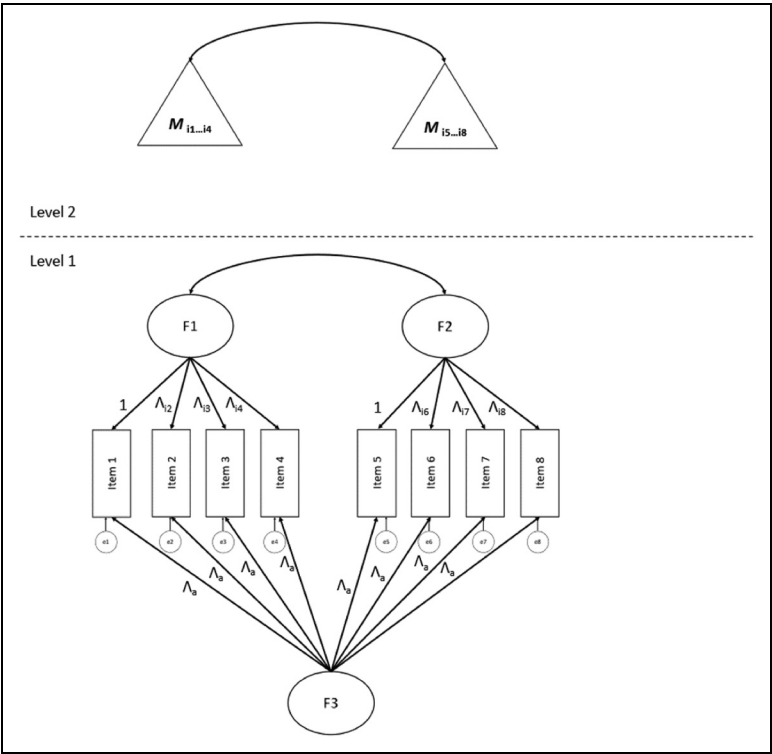


Figure 1. Bi-Factor Model to estimate individual-level effects of an unrelated third factor on nation-level mean relations. *Note:* We estimated a method factor F3, which was conceptually unrelated to the two substantive factors F1 and F2. The loadings of each item on F3 were fixed to be equal across items (varying between $\alpha = 0, .10, .30$ and $.50$ across conditions). The simulated item means per factor were summed ($M_{11...14}$; $M_{15...18}$) and nomological network correlations of the simulated data were computed at level 1 and level 2. See Appendix A for further technical details.

Systematic sources of bias can influence correlations and means at both levels: An empirical observation on support for same-sex marriage

In the previous section, we used simulation to investigate whether systematic bias at the individual level can influence correlations at both individual and aggregate level. We used a simple simulation set-up to demonstrate that validities based on correlations can include systematic bias. Here, we use

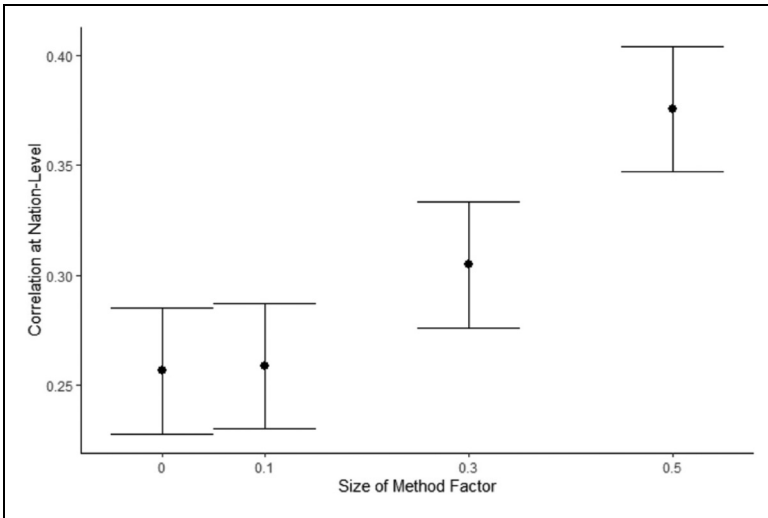


Figure 2. *Simulation showing the effect of systematic bias at individual level on societal level nomological network correlations. Note: True correlation at individual level is .3; Size of Method Factor indicates the loading of observed indicators on the systematic bias factor, Values on the Y-axis indicate the observed correlations at aggregate level (i.e., population level).*

actual data from the World Value Survey to again examine whether systematic variables can influence both means and correlations. We focus on the choice index within the emancipative values index (Welzel, 2014). The index consists of three items, which require individuals to respond whether divorce, abortion and homosexuality are justified (from never (1) to always (10)). Sokolov (2018) recently demonstrated that this three-item index satisfies stringent indicators of invariance² when using a less restrictive approximate Bayesian approach. We want to examine whether there are possible systematic biasing effects even for such a scale where approximate invariance has been demonstrated.

Our departure point is the observation that some of these behaviors are either legal or illegal within specific societies and over time. Because the WVS methodology relies on interviews, it requires individuals to possibly state that something is morally justifiable even if it is illegal in their current context. Of the three behaviors, divorce is probably the least contentious legally and there are relatively few legal restrictions on divorce. The legality of abortion is relatively more complex, with legal norms being influenced by

gestation periods, harm to the mother or child, rape, or social and economic considerations. Among the three issues, same-sex behaviors has the relative clearest legal distinctions and the strictest punishments in case of violation, which includes the possibility of the death penalty in several nations. Here, we do not consider state differences within multistate nations or local laws, but rather focus on national legislation. This approach allows us to examine the effect of legality for which there have been significant changes in some countries over the years. Hence, we can differentiate legal effects from broader societal-level trajectories. Although it is possible to argue that collective behavior (e.g., demonstrations, strikes, voting) can lead to changes in legislation over time (Welzel, 2014), the individual in an interview context does not have an influence on the legal code, but the legal code still may affect responses during the interview. Therefore, the plausible statistical relation is top-down, from institutional nation-level dynamics to individual-level responses during the interview.

We coded the year that same-sex behavior became officially legalized in each nation state (*State-Sponsored Homophobia Report 2019*, 2019). This allows us to test whether legal constraints on same-sex behavior are related to interview responses and whether the legality of the item content has some further effects on the internal and external validity of the choice index (see supplement B for full information). In other words, we tested whether the presence of legal restrictions has an influence on interview responses across nations and within nations over time. In doing so, we make an implicit distinction between personal preferences and external pressures to respond (e.g., avoidance to reveal support for illegal behavior).

First, to what extent does legality of behavior influence the reported justifiability of homosexuality in the interview? We found substantive variation in responses, including variation in responding within nations depending on whether same-sex behavior was officially legalized or not. When probing this effect of legality in multi-level models, we found statistically significant effects of legal restrictions on endorsement of homosexuality within nations. Generally, endorsement for homosexual behavior increased over time, but this effect was significantly dampened by legal restrictions. In other words, there is systematic increase over time for supporting the moral equality of same-sex behavior, but legal restrictions affect expression of opinions in interviews. Responses of individuals in interviews are not free from systematic bias (such as legality of the item content being asked about). Certainly, legal restrictions are an important feature of a society. Our point here is that legal restrictions can have a systematic biasing influence

on the measurement process and the validities of the survey responses. Next, we discuss how this biasing effect impacted both internal and external validities.

Focusing on the internal validities of the choice index, we examined the relationship between the perceived justifiability of homosexuality and both abortion and divorce in a multilevel analysis with random intercepts for both nation and time. We found stronger internal validities in the presence of legal restrictions. This strengthening of validities may be partially driven by restrictions in variation at the individual level, which reduces random variation (e.g., individuals endorsing homosexuality when it is illegal feel very strongly about it) and therefore increases internal validity (the internal emancipation effects around sociosexual liberties becomes more strongly aligned). Note that this pattern is exactly the opposite of the effect claimed by Welzel et al. (2021, see pp. 13ff): an external variable can systematically reduce individual variation which in turn then increases internal validities (instead of decreasing it). Therefore, systematic method effects due to context can strengthen (!) correlations even when variance is being reduced within samples.

Focusing on the external validity of this three-item index, we examined the willingness to fight for one's country as an example item here (see Sokolov, 2018 for a discussion of why this is a suitable item for external validity checks). To test the effects of legality, we ran two sets of multi-level models. First, we ran a multilevel model with random intercepts for nation and time at individual level. Then, we estimated aggregate level models with random intercepts nested in time (which is similar to a nomological network, but accounting for time effects). In both models, we found biasing effects at both individual and nation level. For example, for the nation-level nomological network association, the external validity effect between willingness to fight and choice was only significant when same-sex behavior was legal ($b = -.044$, $se = .008$, $p < .0001$), but not when it was illegal ($b = .005$, $se = .018$, $p = .77$). Systematic bias influences nomological network relations at both individual and nation-level, which requires close attention and careful analysis of possible biasing effects.

Our main point using these data here is that even in the presence of approximate invariance as demonstrated by Sokolov (2018) it is possible to identify systematic impacts of contextual variables that influence the measurement properties and both internal and external validities. To reiterate – correlations with external variables in nomological networks are not free of systematic bias.

Reconceptualizing invariance as an opportunity for conceptual and theoretical maturation and sophistication

The quest to identify bias and invariance in empirical datasets is a fundamental safeguard for making correct inferences about collected data. At the same time, the presence of bias within one's data should be seen as an opportunity for theory development. For example, the careful exploration of invariance tests in value research has demonstrated that economic conditions both within and across nations influence how individuals form cognitive representations of values, which provides important new insights into pathways for societal change (Davidov et al., 2012; Fischer, 2017; Fischer et al., 2011; Fontaine et al., 2008). The observation of variability in invariance parameters provided the opportunity to develop stronger theoretical insights of how societal dynamics are operating. These effects are of direct relevance for the societal changes theorized about by Welzel and colleagues, especially when considering the link between existential conditions and action resources on value transformation. Engaging more carefully with such dynamics can provide exciting new insights into human sociality and cultural dynamics (Fischer, 2017; Gurven, 2018; Lukaszewski et al., 2017). In other words, how people understand and respond to questions may be the key for understanding the potential for societal transformation.

Conclusion

Strong and theoretically expected networks at country level cannot justify the claim that data are free from bias or inequivalence. We used a simple simulation and empirical observations to indicate that systematic biasing effects can have an influence on internal and external validities at nation-level and lead to potentially misleading conclusions. Failing to recognise bias and lack of equivalence can have serious consequences and damage the research field. Discarding one method, such as MGCFA, does not do away with the problem. We recommend a more diversified and principled approach to invariance testing. Bias or inequivalence does form a plausible alternative account for (a part of) the relationships observed at nation level, and this requires unwavering close scrutiny. As a reward, the identification of sources of invariance opens new theoretical avenues worth exploring.

Acknowledgments

We would like to thank Maksim Rudnev for comments on our simulation study and the editor and an anonymous reviewer for comments on a previous version of the manuscript.

Author contributions

All authors contributed to conceptualizing the commentary and the simulation studies; JAK performed the simulation; RF conducted the WVS analyses; all authors revised and agreed on the final version of this commentary.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico, (grant number 310336/2019-8).

ORCID iDs

Ronald Fischer  <https://orcid.org/0000-0002-3055-3955>

Johannes A. Karl  <https://orcid.org/0000-0001-5166-0728>

Johnny Fontaine  <https://orcid.org/0000-0002-5684-0178>

Supplemental Material

The supplemental material for this article is available online.

Notes

1. This point was brought to our attention by a reviewer of a previous draft of this commentary.
2. Sokolov (2018) reported that both the factor loadings and the item intercepts were invariant across samples, satisfying conditions of both metric (equality of factor loadings) and scalar (equality of item intercepts) invariance. These invariance conditions allow direct comparison of latent variables.

References

- Alwin, D. F. 2007. *Margins of Error: A Study of Reliability in Survey Measurement*. Hoboken, NJ: Wiley-Interscience.
- Barrett, P. 2007. "Structural Equation Modelling: Adjudging Model fit." *Personality and Individual Differences* 42(5):815-24. <https://doi.org/10.1016/j.paid.2006.09.018>

- Boer, D., K. Hanke, and J. He. 2018. "On Detecting Systematic Measurement Error in Cross-Cultural Research: A Review and Critical Reflection on Equivalence and Invariance Tests." *Journal of Cross-Cultural Psychology* 49(5):713-34. <https://doi.org/10.1177/0022022117749042>
- Borsboom, D., G. J. Mellenbergh, and J. van Heerden. 2003. "The Theoretical status of Latent Variables." *Psychological Review* 110(2):203-19. <https://doi.org/10.1037/0033-295X.110.2.203>
- Carroll, J. B. 1993. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. pp. ix, 819. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511571312>
- Chen, F. F. 2007. "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance." *Structural Equation Modeling: A Multidisciplinary Journal* 14(3):464-504. <https://doi.org/10.1080/10705510701301834>
- Dalege, J., D. Borsboom, F. van Harreveld, H. van den Berg, M. Conner, and H. L. J. van der Maas. 2016. "Toward a Formalized Account of Attitudes: The Causal Attitude Network (CAN) Model." *Psychological Review* 123(1):2-22. <https://doi.org/10.1037/a0039802>
- Davidov, E., H. Dülmer, E. Schlüter, P. Schmidt, and B. Meuleman. 2012. "Using a Multilevel Structural Equation Modeling Approach to Explain Cross-Cultural Measurement Noninvariance." *Journal of Cross-Cultural Psychology* 43(4):558-75. <https://doi.org/10.1177/0022022112438397>
- Fischer, R. 2017. *Personality, Values, Culture: An Evolutionary Approach*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316091944>
- Fischer, R. and J. A. Karl. 2019. "A Primer to (Cross-Cultural) Multi-Group Invariance Testing Possibilities in R." *Frontiers in Psychology* 10: 1–10. <https://doi.org/10.3389/fpsyg.2019.01507>
- Fischer, R., T. L. Milfont, and V. V. Gouveia. 2011. "Does Social Context Affect Value Structures? Testing the Within-Country Stability of Value Structures With a Functional Theory of Values." *Journal of Cross-Cultural Psychology* 42(2):253-70. <https://doi.org/10.1177/0022022110396888>
- Fontaine, J. R. J., Y. H. Poortinga, L. Delbeke, and S. H. Schwartz. 2008. "Structural Equivalence of the Values Domain Across Cultures: Distinguishing Sampling Fluctuations From Meaningful Variation." *Journal of Cross-Cultural Psychology* 39(4):345-65. <https://doi.org/10.1177/0022022108318112>
- Gurven, M. D. 2018. "Broadening Horizons: Sample Diversity and Socioecological Theory are Essential to the Future of Psychological Science." *Proceedings of the National Academy of Sciences* 115(45):11420-7. <https://doi.org/10.1073/pnas.1720433115>
- Harzing, A.-W. 2006. "Response Styles in Cross-National Survey Research: A 26-Country Study." *International Journal of Cross Cultural Management* 6(2):243-66. <https://doi.org/10.1177/1470595806066332>

- He, J., F. J. R. van de Vijver, A. Dominguez Espinosa, A. Abubakar, R. Dimitrova, B. G. Adams, A. Aydinli, K. Atitsogbe, I. Alonso-Arbiol, M. Bobowik, R. Fischer, V. Jordanov, S. Mastrotheodoros, F. Neto, Y. J. Ponizovsky, J. Reb, S. Sim, L. Sovet, and D. Stefenel, ... A. Villieux. 2015. "Socially Desirable Responding: Enhancement and Denial in 20 Countries." *Cross-Cultural Research* 49(3):227-49. <https://doi.org/10.1177/1069397114552781>
- Kang, Y., D. M. McNeish, and G. R. Hancock. 2016. "The Role of Measurement Quality on Practical Guidelines for Assessing Measurement and Structural Invariance." *Educational and Psychological Measurement* 76(4):533-61. <https://doi.org/10.1177/0013164415603764>
- Lukaszewski, A. W., M. Gurven, C. R. von Rueden, and D. P. Schmitt. 2017. "What Explains Personality Covariation? A Test of the Socioecological Complexity Hypothesis." *Social Psychological and Personality Science* 8(8):943-52. <https://doi.org/10.1177/1948550617697175>
- Martinez Gutierrez, N. and R. Cribbie. (2021). "Incidence and Interpretation of Statistical Suppression in Psychological Research." *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement* 53(4):480-488. <https://doi.org/10.1037/cbs0000267>
- Malda, M., F. J. R. van de Vijver, and Q. M. Temane. 2010. "Rugby Versus Soccer in South Africa: Content Familiarity Contributes to Cross-Cultural Differences in Cognitive Test Scores." *Intelligence* 38(6):582-95. <https://doi.org/10.1016/j.intell.2010.07.004>
- Marsh, H. W., K.-T. Hau, and Z. Wen. 2004. "In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings." *Structural Equation Modeling: A Multidisciplinary Journal* 11(3):320-41. https://doi.org/10.1207/s15328007sem1103_2
- Matsumoto, D. R. and F. J. R. van de Vijver (Eds.). 2011. *Cross-cultural Research Methods in Psychology*. Cambridge: Cambridge University Press.
- Messick, S. 1981. "Constructs and Their Vicissitudes in Educational and Psychological Measurement." *Psychological Bulletin* 89(3):575-88. <https://doi.org/10.1037/0033-2909.89.3.575>
- Messick, S. 1989. "Validity." Pp. 13-103 in *Educational Measurement*, 3rd ed. Cambridge: American Council on Education.
- Nye, C. D. and F. Drasgow. 2011. "Effect Size Indices for Analyses of Measurement Equivalence: Understanding the Practical Importance of Differences Between Groups." *Journal of Applied Psychology* 96(5):966-80. <https://doi.org/10.1037/a0022955>

- Rodriguez, A., S. P. Reise, and M. G. Haviland. 2016. "Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures." *Journal of Personality Assessment* 98(3):223-37. <https://doi.org/10.1080/00223891.2015.1089249>
- Schmidt, F. L. and J. E. Hunter. 1998. "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings." *Psychological Bulletin* 124(2):262-74. <https://doi.org/10.1037/0033-2909.124.2.262>
- Scribner, S. and M. Cole. 1981. *The Psychology of Literacy*. Cambridge, Mass: Harvard University Press. <https://doi.org/10.4159/harvard.9780674433014>
- Serpell, R. 1979. "How Specific are Perceptual Skills? A Cross-Cultural Study of Pattern Reproduction." *British Journal of Psychology* 70(3):365-80. <https://doi.org/10.1111/j.2044-8295.1979.tb01706.x>
- Smith, P. B. 2004. "Acquiescent Response Bias as an Aspect of Cultural Communication Style." *Journal of Cross-Cultural Psychology* 35(1):50-61. <https://doi.org/10.1177/0022022103260380>
- Smith, P. B., V. L. Vignoles, M. Becker, E. Owe, M. J. Easterbrook, R. Brown, D. Bourguignon, R. B. Garðarsdóttir, R. Kreuzbauer, B. C. Ayala, M. Yuki, J. Zhang, S. Lv, P. Chobthamkit, J. L. Jaafar, R. Fischer, T. L. Milfont, A. Gavreliuc, and P. Baguma, ... C. Harb. 2016. "Individual and Culture-Level Components of Survey Response Styles: A Multi-Level Analysis Using Cultural Models of Selfhood." *International Journal of Psychology* 51(6):453-63. <https://doi.org/10.1002/ijop.12293>
- Sokolov, B. 2018. "The Index of Emancipative Values: Measurement Model Misspecifications." *American Political Science Review* 112(2):395-408. <https://doi.org/10.1017/S0003055417000624>
- State-Sponsored Homophobia report 2019: Global legislation overview update. 2019, December 10. ILGA. <https://ilga.org/state-sponsored-homophobia-report-2019-global-legislation-overview>.
- van de Vijver, F. J. R. 2011. "Bias and Real Differences in Cross-Cultural Differences: Neither Friends nor Foes." Pp. 235-57 in *Fundamental Questions in Cross-Cultural Psychology*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511974090.011>
- van de Vijver, F. J. R. and K. Leung. 1997. *Methods and Data Analysis for Cross-Cultural Research / Fons van de Vijver, Kwok Leung*. Thousand Oaks: Sage Publications.
- Van De Vijver, F. J. R. and Y. H. Poortinga. 1982. "Cross-Cultural Generalization and Universality." *Journal of Cross-Cultural Psychology* 13(4):387-408. <https://doi.org/10.1177/0022002182013004001>

- van de Vijver, F. J. R. and Y. H. Poortinga. 2002. "Structural Equivalence in Multilevel Research." *Journal of Cross-Cultural Psychology* 33(2):141-56. <https://doi.org/10.1177/0022022102033002002>
- van Herk, H., Y. H. Poortinga, and T. M. M. Verhallen. 2004. "Response Styles in Rating Scales: Evidence of Method Bias in Data From Six EU Countries." *Journal of Cross-Cultural Psychology* 35(3):346-60. <https://doi.org/10.1177/0022022104264126>
- Van De Vijver, F. J. R., D. A. van Hemert, and Y. H. Poortinga, (Eds.). 2008. *Multilevel Analysis of Individuals and Cultures*. New York: Lawrence Erlbaum Associates.
- Weijters, B., M. Geuens, and H. Baumgartner. 2013. "The Effect of Familiarity with the Response Category Labels on Item Response to Likert Scales." *Journal of Consumer Research* 40(2):368-81. <https://doi.org/10.1086/670394>
- Welkenhuysen-Gybels, J., J. Billiet, and B. Cambré. 2003. "Adjustment for Acquiescence in the Assessment of the Construct Equivalence of Likert-Type Score Items." *Journal of Cross-Cultural Psychology* 34(6):702-22. <https://doi.org/10.1177/0022022103257070>
- Welzel, C. 2014. *Freedom Rising: Human Empowerment and the Quest for Emancipation*. New York: Cambridge University Press.
- Welzel, C., L. Brunkert, S. Kruse, and R. F. Inglehart. 2021. "Non-invariance? An Overstated Problem With Misconceived Causes." *Sociological Methods & Research* Online first publication, 1–33. <https://doi.org/10.1177/0049124121995521>

Author Biographies

Ronald Fischer is Professor in Psychology at Victoria University of Wellington, New Zealand and Researcher in the Neuroscience Unit at D'Or Institute for Research & Teaching, Brazil. He is a Fellow of the Royal Society NZ Te Aparangi and the Association for Psychological Science. His research focuses on cultural and evolutionary multilevel dynamics of values, beliefs, personality and wellbeing.

Johannes Alfons Karl is a Lecturer in Psychology at the Victoria University of Wellington, New Zealand. He is currently an Associate Editor for Mindfulness. His work focuses on psychometric properties of tests to examine individual differences and motivational goals, with a special interest in effects of ecological disturbances on psychometric properties (e.g., COVID-19 effects on measurement properties).

Johnny R. J. Fontaine is professor at the faculty of Psychology and Educational Sciences of Ghent University in Belgium. He teaches cross-cultural psychology and psychological assessment. He is the past president of the European Association of Psychological Assessment and an associate editor of *Journal of Cross-Cultural*

Psychology. His research interests are in the domains of cross-cultural assessment, emotions, and values.

Ype H. Poortinga is emeritus professor of cross-cultural psychology at Tilburg University in the Netherlands and at the Catholic University of Leuven in Belgium. His publications reflect a long-term interest in the valid comparison of data from distinct populations. He is an honorary fellow and former president of the International Association for Cross-Cultural Psychology.