This (method) is (not) fine

Wicherts, Jelte M

[Link to publication in Tilburg University Research Portal](#)

Debate

THIS (METHOD) IS (NOT) FINE

JELTE M. WICHERTS[1]

Department of Methodology and Statistics, Tilburg University, Tilburg, the Netherlands

Summary. In their response to my criticism of their recent article in *Journal of Biosocial Science* (te Nijenhuis *et al*., 2017), te Nijenhuis and van den Hoek (2018) raise four points none of which concerns my main point that the method of correlated vectors (MCV) applied to item-level data represents a flawed method. Here, I discuss te Nijenhuis and van den Hoek's four points. First, I argue that my previous application of MCV to item-level data showed that the method can yield nonsensical results. Second, I note that meta-analytic corrections for sampling error, imperfect measures, restriction of range and unreliability of the vectors are futile and cannot help fix the method. Third, I note that even with perfect data, the method can yield negative correlations. Fourth, I highlight the irrelevance of te Nijenhuis and van den Hoek (2018)'s point that my comment had not been published in a peer-reviewed journal by referring to my articles in 2009 and 2017 on MCV in peer-reviewed journals.

In their response to my criticism of their recent article in *Journal of Biosocial Science* (te Nijenhuis *et al*., 2017), te Nijenhuis and van den Hoek (2018; hereafter TeNvdH) raise four points, none of which concerns my main point that the method of correlated vectors (MCV) applied to item-level data represents a flawed method. Their response highlights their continued reluctance to confront the psychometric problems with MCV at the item level. I will discuss their four points in turn.

First, TeNvdH indicate that I used MCV at the item level in my own work with colleagues (Kan *et al*., 2013). But as anyone could readily glean from the abstract, Kan *et al*.'s (2013) application of MCV revolved around subtests rather than dichotomous item scores. TeNvdH appear to confuse (approximately) normal data from subtests as

---

[1] Email: j.m.wicherts@uvt.nl

modelled in the context of linear factor analysis with data that are dichotomous (i.e. scored 0 or 1) and modelled as appropriately non-linear in the context of item response theory. I had hoped that TeNvdH would finally discuss my article on MCV applied to the item level (Wicherts, 2017), in which I certainly applied MCV to item-level data to illustrate its problems empirically. The empirical application of MCV at the item level in Wicherts (2017) was inspired by te Nijenhuis' scepticism of data simulations, so I am disappointed that TeNvdH fail to discuss the nonsensical results I found with this method in real data comparing entirely unrelated measures of intelligence, anxiety, anger and the big five personality traits.

Second, TeNvdH refer to meta-analytic corrections developed by Hunter and Schmidt (2004), and rightly indicated that I did not discuss these corrections. The reason again is that my main criticism (Wicherts, 2017) is levelled against the use of MCV at the item level, for which the Schmidt and Hunter corrections were not developed. TeNvdH appear of the opinion that it is fine to apply a statistical method developed for one type of data (continuous) to a completely different type of data (dichotomous) without first trying to understand (or even discuss) its statistical complexities. Crucially, my psychometric argument was based on a scenario in which the items measured g perfectly, the loadings were not subject to any error, the loadings covered the entire range and there was no sampling error. In these cases, corrections for sampling error, reliabilities of the vectors, imperfect measures of g, and restriction of range in the loadings are completely futile. These corrections do not magically unbend the non-linearity implicit in MCV relations at the item level, even with perfect data.

Third, TeNvdH refer to an early criticism of MCV at the scale level by Schönemann (1997) that they consider to be 'eerily similar' to my criticism of MCV at the item level. Again, Schöneman's discussion related to linear data that has little bearing on my point that MCV yields complex non-linear relations when applied to dichotomous items. Such relations might certainly also create negative relations, as I clearly indicated in my 2017 article: 'the variation in MCV correlations is due to the complex non-linearity that is caused by the restricted range of the relevant CTT item statistics (that go from -1 to 1) and their inherent group-specificity. That MCV correlations based on item scores can be very low (negative even) in the population even if Spearman's hypothesis were true highlights that MCV lacks sensitivity: Even if Spearman were right and the test measures g and nothing else, MCV correlations close to 1 will only occur in rare cases' (Wicherts, 2017, p. 30). I provided an Excel file as supplementary file that allows anyone to assess this readily. The vectors can correlate negatively and positively in cases that align perfectly with g theory, and they can positively correlate highly even when the data are nonsensical in practice (Wicherts, 2017). The correlations can be negative when one uses loadings that are estimated appropriately in a latent variable framework as in the example described by TeNvdH.

Fourth, TeNvdH suggest that I misused my position as reviewer by referring to an unpublished paper when I reviewed their work. They perhaps forgot that this particular paper (Wicherts, 2017) spent a lot of time in the peer review pipeline at *Intelligence* because it required summary data that te Nijenhuis *et al.* had not reported in their paper. My paper was considerably delayed because it took te Nijenhuis and colleagues a full year to share the needed data to complete my analyses. In earlier correspondence,

te Nijenhuis explicitly refused to read an earlier version of Wicherts (2017) because it had not yet been published in a peer-reviewed journal, and TeNvdH now repeat this argument as though this is legal case wherin certain evidence is not allowed to be considered. The paper has been published in the peer-reviewed literature in late 2016, so TeNvdH can now safely read it. Besides, they could have also considered my earlier peer-reviewed work (Wicherts & Johnson, 2009) that made a very similar case against the use of MCV at the item level.

References

Hunter, J. E. & Schmidt, F. L. (2004) *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, 2nd edition. Sage, London.

Kan, K. J., Wicherts, J. M., Dolan, C. V. & Van Der Maas, H. L. (2013) On the nature and nurture of intelligence and specific cognitive abilities: the more heritable, the more culture dependent. *Psychological Science* **24**(12), 2420–2428.

Schönemann, P. H. (1997) Famous artifacts: Spearman's hypothesis. *Cahiers de Psychologie Cognitive* **16**, 665–694.

te Nijenhuis, J., Batterjee, A. A., van den Hoek, M., Allik, J. & Sukhanovskiy, V. (2017) Spearman's hypothesis testing comparing Saudi Arabian children and adolescents with various other groups of children and adolescents on the items of the Standard Progressive Matrices. *Journal of Biosocial Science* **49**, 634–647.

te Nijenhuis, J. & van den Hoek, M. (2018) Analysing group differences in intelligence using the psychometric meta-analytic-method of correlated vectors hybrid model: a reply to Wicherts (2018) attacking a strawman. *Journal of Biosocial Science*, doi:10.1017/S0021932018000160.

Wicherts, J. M. (2017) Psychometric problems with the method of correlated vectors applied to item scores (including some nonsensical results). *Intelligence* **60**, 26–38.

Wicherts, J. M. & Johnson, W. (2009) Group differences in the heritability of items and test scores. *Proceedings of the Royal Society: Series B* **276**, 2675–2683.