Comment

Intelligence Is What the Intelligence Test Measures. Seriously

Han L. J. van der Maas [1,*], Kees-Jan Kan [2] and Denny Borsboom [1]

[1] Psychological Methods, Department of Psychology, University of Amsterdam, Weesperplein 4, room 207, Amsterdam 1018 NX, The Netherlands; E-Mail: D.Borsboom@uva.nl (D.B.)

[2] Department of Biological Psychology, VU University, van der Boechorststraat 1, Amsterdam 1081 BT, The Netherlands; E-Mail: k.j.kan@vu.nl (K.-J.K.)

* Author to whom correspondence should be addressed; E-Mail: h.l.j.vandermaas@uva.nl; Tel.: +31-20-525-6678.

Abstract: The mutualism model, an alternative for the *g*-factor model of intelligence, implies a formative measurement model in which "*g*" is an index variable without a causal role. If this model is accurate, the search for a genetic of brain instantiation of "*g*" is deemed useless. This also implies that the (weighted) sum score of items of an intelligence test is just what it is: a weighted sum score. Preference for one index above the other is a pragmatic issue that rests mainly on predictive value.

Keywords: mutualism model; formative model; intelligence test; *g*-factor; IQ

The first thing psychology students learn about intelligence is that Boring's 1923 definition of intelligence as what an IQ-test measures is just silly [1]. In line with the literature on the topic, both Johnson [2] and Hunt & Jaeggi [3] do not take it seriously. However, given recent developments in the theory of intelligence, we think that there is reason to reconsider our opinion on this topic.

Empirically, the core of intelligence research rests on the positive manifold: the fact that all intelligence subtests, ranging from scholastic tests to tests of social intelligence, correlate positively. This correlational pattern can be modeled statistically with principal components analysis or with factor models. Although principal components analysis and factor analysis are sometimes considered to be more or less interchangeable, they are very different models [4]. The factor model is a reflective latent variable model, in which the factor is a hypothesized entity that is posited to provide a putative

explanation for the positive manifold [5]. The principal components model is a formative model, in which the components are conveniently weighted total scores; these are composites of the observed data, which do not provide an explanation of the positive manifold, but rather inherit their structure entirely from the data [6].

Thus, the factor model embodies the idea that there is a common cause "out there" that we "detect" using factor analysis, and that should have an independently ascertainable identity in the form of, say, a variable defined on some biological substrate [7]. The principal components model does not say anything about the nature of the correlations in the positive manifold. It does not posit testable restrictions on the data, and therefore is better thought of as a data reduction model than as a explanatory model. Importantly, in formative models, the nature of the constructed components is fixed by the subtests used to determine them: a different choice of subtests yield conceptually different components (even though these may be highly correlated; see also [8]). In contrast, the latent variable in the factor model is not specifically tied to any set of subtests: if the model is true, the latent variable can in principle be measured by any suitable set of indicators that depends on it and fulfills relevant model requirements. Although different choices of such indicators may change the precision of the measurements, they need not change the nature of the latent variable measured.

Clearly, the classical *g* model, as for instance discussed by Jensen [9] is a reflective model: whatever *g* is, it is thought to *explain* scores on tests, correlations between tests, and individual differences between subjects or groups of subjects. In other words, the *g*-factor is posited as the *common cause* of the correlations in the positive manifold. Recently, however, an alternative explanation for the positive manifold has been proposed in the form of the mutualism model [10]. In this model, the correlations between test scores are not explained through the dependence on a common latent variable, but as a result of reciprocal positive interactions between abilities and processes that play key roles in cognitive development, like memory, spatial ability, and language skills. The model explains key findings in intelligence research, such as the hierarchical factor structure of intelligence, the low predictability of intelligence from early childhood performance, the age differentiation effect, the increase in heritability of *g*, and is consistent with current explanations of the Flynn effect [10].

It is interesting to inquire what the status of *g* should be if such a mutualism model were true. Of course, in this situation, one does not measure a common latent variable through IQ-tests, for there is no such latent variable. Rather, the mutualism model would support a typical formative model [11]. Such a formative model is also implied by a much older alternative for the *g* model, sampling theory [12]. In a formative model, the factor score estimates that results from applications of *g* factor models represent just another weighted sum of test scores and should be interpreted as index statistics instead of as a latent variable. Index statistics, such as the Dow Jones Industrial Average, the Ocean Health index and physical health indexes, evidently do not cause economic growth or healthy behaviors. Instead, they result from or supervene on them [6].

Traditionally, the principal components model has been seen as the weak sister of the factor model, which was thought to give the better approach to modeling IQ subtest scores [13]. However, under the mutualism model, the situation is reversed. The principal components model in fact yields as good a composite as any other model. The use of the factor model, in this view, amounts to cutting butter with a razor: it represents and extraordinarily complicated and roundabout way of constructing composite

scores that are in fact no different from principal component scores. In particular, factor score estimates do not yield measurements of a latent variable that leads an independent life in the form a biological substrate or such. They are just weighted sum scores.

Thus, the mutualism model explains the positive manifold but at the same time denies the existence of a realistic *g*. As a result, it motivates a formative interpretation of the factor analytic model. This has many implications. First, if *g* is not a causal source of the positive manifold, the search for a gene or brain area "for *g*" will be fruitless [14]. Again, the comparison with health is instructive. There are no specific genes "for health", and health has no specific location in the body. Note that this line of reasoning does not apply to genetic and brain research on *components* of intelligence (for instance working memory) as these components often do have a realistic reflective interpretation. Working memory capacity may very well be based on specific and independently identifiable brain processes, even if *g* is not.

The implications of a mutualism model for approaches to measurement are likewise significant [4,14,15]. One crucial difference between reflective and formative models, for instance, concerns the role of the indicator variables (items or subtests). As noted above, in a reflective model these indicators are exchangeable. Therefore, different working memory tests, with different factor loadings, could be added to a test battery without changing the nature of the measured *g* factor. Also, measurements of *g* can be better improved by simply adding more and more relevant tests. Tests can also be ordered in how well they measure *g*, for instance by looking at patterns of factor loadings or by computing indices of measurement precision.

In formative models, however, indicators are not exchangeable, unless they are completely equivalent. There is no universally optimal way to compute the composite scores that function as an index; instead, this choice rests on pragmatic grounds. For example, the justification for the choice of index may lie in its predictive value. The best test is simply the test that best predicts educational, societal or job success. The choice of indicators may however also depend on our "cognitive" environment. When a certain cognitive capacity, say computational thinking, is valued to a greater extent in the current society, intelligence tests may be adapted to reflect that. In an extended mutualistic model, in which reciprocal effects take place via the environment (a gene-environment model; see [16,17]), intelligence testing could even be extended to the assessment of features of the environment that play a positively reinforcing role in promoting the mutualistic processes that produce the positive manifold. On this viewpoint, the number of books somebody owns might very well be included in the construction of composites under an index model of intelligence.

Finally, in a formative interpretation of IQ test scores there really is no such thing as a separate latent variable that we could honor with the term "intelligence", and it is questionable whether one should in fact use the word "intelligence measurement" at all in such a situation [18]. However, if one insists on keeping the terminology of measurement around, there is little choice except to bite the bullet: Interpreted as an index, intelligence is whatever IQ-tests measure. Seriously.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Boring, E.G. Intelligence as the tests test it. *New Repub.* **1923**, *36*, 35–37.
2. Johnson, W. Whither intelligence research? *J. Intell.* **2013**, *1*, 25–35.
3. Hunt, E.; Jaeggi, S.M. Challenges for research on intelligence. *J. Intell.* **2013**, *1*, 36–54.
4. Edwards, J.R.; Bagozzi, R.P. On the nature and direction of relationships between constructs and measures. *Psychol. Methods* **2000**, *5*, 155–174.
5. Haig, B.D. Exploratory factor analysis, theory generation, and scientific method. *Multivar. Behav. Res.* **2005**, *40*, 303–329.
6. Markus, K.; Borsboom, D. *Frontiers of Validity Theory: Measurement, Causation, and Meaning*; Routledge: New York, NY, USA, 2013.
7. Kievit, R.A.; Romeijn, J.W.; Waldorp, L.J.; Wicherts, J.M.; Scholte, H.S.; Borsboom, D. Mind the gap: A psychometric approach to the reduction problem. *Psychol. Inq.* **2011**, *22*, 1–21.
8. Howell, R.D.; Breivik, E.; Wilcox, J.B. Is formative measurement really measurement? *Psychol. Methods* **2007**, *12*, 238–245.
9. Jensen, A.R. *The g Factor: The Science of Mental Ability*; Praeger: Westport, CT, USA, 1999.
10. Van der Maas, H.L.J.; Dolan, C.V.; Grasman, R.P.P.P.; Wicherts, J.M.; Huizenga, H.M.; Raijmakers, M.E.J. A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychol. Rev.* **2006**, *113*, 842–861.
11. Schmittmann, V.D.; Cramer, A.O.; Waldorp, L.J.; Epskamp, S.; Kievit, R.A.; Borsboom, D. Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas Psychol.* **2013**, *31*, 43–53.
12. Thomson, G.H. A hierarchy without a general factor. *Br. J. Psychol.* **1916**, *8*, 271–281.
13. Bartholomew, D.J. *Measuring Intelligence: Facts and Fallacies*; Cambridge University Press: Cambridge, UK, 2004.
14. Chabris, C.F.; Hebert, B.M.; Benjamin, D.J.; Beauchamp, J.; Cesarini, D.; van der Loos, M.; Laibson, D. Most reported genetic associations with general intelligence are probably false positives. *Psychol. Sci.* **2012**, *23*, 1314–1323.
15. Bollen, K.A.; Lennox, R. Conventional wisdom on measurement: A structural equation perspective. *Psychol. Bull.* **1991**, *110*, 305–314.
16. Dickens, W.T.; Flynn, J.R. Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychol. Rev.* **2001**, *108*, 346–369.
17. Kan, K.J.; Wicherts, J.M.; Dolan, C.V.; van der Maas, H.L.J. On the nature and nurture of intelligence and specific cognitive abilities the more heritable, the more culture dependent. *Psychol. Sci.* **2013**, *24*, 2420–2428.
18. Howell, R.D.; Breivik, E.; Wilcox, J.B. Is formative measurement really measurement? *Psychol. Methods* **2007**, *12*, 238–245.