Intelligence 73 (2019) 52–62

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Extending psychometric network analysis: Empirical evidence against *g* in favor of mutualism?

Kees-Jan Kan[a,*], Han L.J. van der Maas[b], Stephen Z. Levine[c]

[a] *Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, the Netherlands*
[b] *Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands*
[c] *Department of Community Mental Health, University of Haifa, Haifa, Israel*

ABSTRACT

The current study implements psychometric network analysis within the framework of confirmatory (structural equation) modeling. Utility is demonstrated by three applications on independent data sets. The first application uses WAIS data and shows that the same kind of fit statistics can be produced for network models as for traditional confirmatory factor models. This can assist deciding between factor analytical and network theories of intelligence, e.g. *g* theory versus mutualism theory. The second application uses the 'Holzinger and Swineford data' and illustrates how to cross-validate a network. The third application concerns a multigroup analysis on scores on the Brief Test of Adult Cognition by Telephone (BCATC). It exemplifies how to test if network parameters have the same values across groups. Of theoretical interest is that in all applications psychometric network models outperformed previously established (*g*) factor models. Simulations showed that this was unlikely due to overparameterization. Thus the overall results were more consistent with mutualism theory than with mainstream *g* theory. The presence of common (e.g. genetic) influences is not excluded, however.

1. Introduction

The *description* of the structure of individual differences in cognitive performance and the *explanation* of the etiology of these differences are two major themes within differential psychology. Both are closely connected to advances in statistics and psychometrics (c.f., Jensen, 1998; Kovacs & Conway, 2016; Mackintosh, 2011; van der Maas, Kan, Marsman, & Stevenson, 2017). The first factor analysis, for instance, was carried out on scholastic achievement data and concerned an investigation of the correlational structure among individual differences in subjects as varied as Classics, French, English, Math, Pitch, and Music (Spearman, 1904). According to the (single) common factor model that Spearman devised, the variance in each of these cognitive measures could be *described* as partly shared among all observed variables and as partly unique to the measure (see Fig. 1a). Spearman's *explanation* of this structure was that all observed variables (i.e. the subjects' scores) were influenced by two kinds of unobserved sources (causes) of variance: (1) a single (unitary) source of variance with common effects, dubbed '*g*', which stands for general intelligence or general cognitive ability, and (2) a variety of sources of variance that were unique to each observed variable.

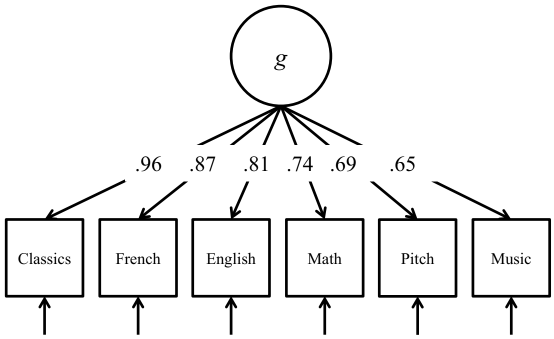
Once broader sets of cognitive variables were analyzed, the single common factor model transpired as unsatisfactory, due to the presence of residual positive correlations among the observed variables (Burt, 1909; Burt, 1911). Consequently, in order to accommodate multiple factors of intelligence, more elaborated factor analytic techniques were developed (Burt, 1909, 1911; Holzinger & Swineford, 1937; Schmid & Leiman, 1957; Thurstone, 1931). The development of orthogonal and oblique rotation (Thurstone, 1947) highlighted that the hypothesis of multiple independent common sources of intelligence was also problematic, as orthogonal model-implied correlation structures did not match observed correlation structures either. Oblique first order factor models, in which factors are allowed to correlate, provided better matches. Next, in order to explain the structure among the correlations among those factors, higher order factor modeling was introduced (Schmid & Leiman, 1957). In addition, confirmatory techniques were developed (Jöreskog, 1967; Jöreskog & Sörbom, 1989), which provided the opportunity to compare factor models statistically and so also the means to distinguish between competing factor theoretical explanations of the correlation structures.

Whereas Spearman's single common factor model of intelligence was relatively soon discarded as providing a satisfactory description of the structure of intelligence (Burt, 1909), the advances in psychometric modeling retained Spearman's hypothesis of an underlying general factor of intelligence, so that more than a century later '*g*' still plays a major role in mainstream psychometric theories of human psychometric intelligence (Gottfredson, 2016). Indeed, expert consensus dictates the "given" that the structure of psychometric intelligence can be

* Corresponding author at: Research Institute of Child Development and Education, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS, Amsterdam, the Netherlands.

E-mail address: k.j.kan@uva.nl (K.-J. Kan).

https://doi.org/10.1016/j.intell.2018.12.004
Received 16 January 2018; Received in revised form 14 June 2018
Available online 29 January 2019



Example Mutualism Matrix **M**

	Classics	French	English	Math	Pitch	Music
Classics	0.00	0.28	0.21	0.17	0.14	0.13
French	0.28	0.00	0.15	0.13	0.11	0.10
English	0.21	0.15	0.00	0.11	0.10	0.09
Math	0.17	0.13	0.11	0.00	0.09	0.08
Pitch	0.14	0.11	0.10	0.09	0.00	0.07
Music	0.13	0.10	0.09	0.08	0.07	0.00
Residual variance	0.21	0.34	0.44	0.54	0.61	0.65

Fig. 1. (a) Spearman's (1904) single common factor model (left panel), and (b) a mutualism model, **M**, that predicts the exact same correlation matrix (right panel). Note: Both concern solutions obtained in OpenMx using Maximum Likelihood. Matrix **M** contains regression weights, see main text.

arranged hierarchically, with a general factor at the apex and less broad factors arrayed below (Carroll, 2003; Jensen, 1998; Johnson, Bouchard Jr, Krueger, McGue, & Gottesman, 2004; Neisser et al., 1996; Schneider & McGrew, 2018). This interpretation is also given to the bi-factor models in which higher order factor models are nested, but which some prefer over those higher order models (e.g., Gignac & Watkins, 2013). We note that from a *descriptive* (statistical) point of view bi-factor models may tend to fit better, but also that from an *explanatory* (substantive theoretical) perspective, a bifactor model of intelligence is considered unsatisfactory (e.g., Jensen, 1998; Hood, 2008). Decisions as to which model to adopt as a the best model should rely on both fit and theory, not on fit itself (Morgan, Hodge, Wells, & Watkins, 2015; Murray & Johnson, 2013). In other words, theory drives, fit assists.

Possibly this is one reason that the substantial interpretation of the general factor of intelligence remains unsettled (Carroll, 2003; Horn, 1998; Nisbett et al., 2012; Schneider & McGrew, 2018), hence despite the wide acceptance of the possibility to extract such a factor. Indeed, the nature of the general factor persists as a topic of heated debate, and one with renewed attention (Bartholomew, Deary, & Lawn, 2009; Kovacs & Conway, 2016). This debate may be summarized as follows. Firstly, provided the general factor represents indeed a single, unitary *source* (*cause*) of common variance, the question remains which one. Is *g* ‘mental energy’, as Spearman (1927) himself suggested? Or is it, for instance, abstract reasoning ability (Gustafsson, 1984), working memory capacity (Kyllonen & Christal, 1990), neural plasticity (Garlick, 2002), or the efficiency of mitochondrial functioning (Geary, 2018)? Secondly, that the general factor represents a single, unitary source of variance is not a given, but a hypothesis, which – like any other scientific hypothesis – requires empirical scrutiny.

In this regard, the results of other (non-psychometric) lines of research are of importance. These include, among others, educational science, sociology, (neuro)biology, and genetics. Also within these lines statistical and psychometric modeling has added value, as the following examples clarify. Cortical thickness and (total, grey, and white matter) volume are well-known neurobiological correlates with intelligence (for an overview of biological correlates, see Jensen, 1998), but with help of more advanced (structural equation) modeling (SEM) one can test competing hypotheses about the way such variables may have impact (Kievit et al., 2012). Similarly, family studies and later molecular genetic studies have shown that IQ and *g* are substantially heritable (see Plomin & von Stumm, 2018, for a recent overview) and that the same holds for the many brain correlates of intelligence (Hill et al., 2018; Plomin & von Stumm, 2018). Again with the help of SEM, one can test competing hypotheses about how genetic and environmental influences may have their effects (Neale & Cardon, 1994). It allows one to investigate, for instance, if genetic and environmental influences are mediated by a latent factor (e.g., *g*) or that they more likely follow independent pathways. Such research is all but trivial, because violation of assumptions concerning the etiology of the phenotype decreases the power to detect genetic variants (van der Sluis, Posthuma, & Dolan, 2013). Genetics and statistical modeling can thus inform each other (Franić et al., 2013; Grotzinger et al., 2018). More generally, both psychometric and non-psychometric lines of research are necessary to understand the concept of intelligence and their integration is welcomed.

The essence of statistical modeling and model selection (Kline, 2015) is the combination of Popperian logic (Popper, 2005) and Occam's razor or ‘the law of parsimony’. That is, hypotheses can never be proven, but they can be disproven and competing hypotheses can be rejected for being too complex or too simplistic (all else equal). This may thus favor the one hypothesis over the other. In this regard, outcomes of recent studies using bivariate latent change models (McArdle & Hamagami, 2001) on cognitive developmental data (Hofman et al., 2018; Kievit et al., 2017) are relevant to the debate concerning the status of *g*, as they favored the so-called mutualism model of intelligence (van der Maas et al., 2006; see discussion section) over the *g* hypothesis. In the words of Kievit et al., 2017:

> [A] mutualism model, which proposes that basic cognitive abilities directly and positively interact during development, provides the best account of developmental changes. Individuals with higher scores in vocabulary showed greater gains in matrix reasoning and vice versa. These dynamic coupling pathways are not predicted by other accounts and provide a novel mechanistic window into cognitive development. (p. 1).

In brief, the mutualism model of intelligence is a model of cognitive development that was inspired by research in ecosystem modeling, where the dynamics between variables are due to reciprocal causation. The key idea is that such reciprocal causation also occurs among cognitive abilities during their development.

As a formal model, the mutualism model describes intraindividual change (growth) in a given cognitive ability as a function of both (1) autonomous growth, i.e., growth that does not dependent upon other cognitive abilities, and (2) growth due to the influence of the development of other cognitive abilities. The autonomous function is assumed to follow a logistic shape (although this assumption is not crucial). This shape reflects, firstly, the idea of rapid initial growth, and, secondly, that of eventual slowing down due to the presence of limiting capacities. Unlike in *g* theory, these capacities are (or can be) considered statistically independent. Yet, because the growth of a given cognitive ability is not only limited by its own, specific limiting capacity, but is also affected by the level of other cognitive abilities (through the dynamical interactions), and thus by their corresponding limiting capacities, the cognitive abilities themselves become positively correlated throughout the course of their development.

Mutualism thus provides an alternative explanation of the positive manifold, i.e. the robust finding and Spearman's original observation that cognitive performance measures correlate positively with another. The strength of the mutualistic interactions between cognitive abilities can be represented in an interaction matrix, **M** (see Fig. 1b, for an example). If these strengths are assumed equal over individuals, and provided the systems near their equilibrium, the values in this matrix represent regression weights of the direct influences of cognitive abilities on each other (van der Maas et al., 2006; see also Appendix).

As has been noted (van der Maas et al., 2017), mutualism – the idea of dynamic coupling between cognitive abilities – aligns neatly with some of the latest and most rapid developments in psychometrics, namely psychometric network modeling (Borsboom, 2008; Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012). To date, psychometric network analyses have been carried out on indicators of depression (Madhoo & Levine, 2016; van Borkulo et al., 2015), schizophrenia (Levine & Leucht, 2016; van Rooijen et al., 2018), autism (Ruzzano, Borsboom, & Geurts, 2015), loneliness (Fried et al., 2015), and many more constructs related to psychopathology (for a review, see Fried et al., 2017). Although it was psychopathology that has motivated this advancement (Borsboom, 2008; Borsboom & Cramer, 2013), this type of analysis is gaining popularity in other fields as well, including personality (Costantini et al., 2015), attitudes (Dalege, Borsboom, Harreveld, & van der Maas, 2017), interests (Sachisthal et al., 2018), and also cognition and intelligence (Golino & Demetriou, 2017; van der Maas et al., 2017).

The key differences between traditional factor analysis and psychometric network analysis once applied to cognitive data are summarized in Table 1. It shows that one may conceptualize cognitive abilities as being related to each other directly, rather than through common, unobserved variables on which they depend. Indeed, the connections between any pairs of cognitive variables can be modeled using (full or full partial) correlations only, hence without postulating any latent factors. In a psychometric network model, the extent to which a specific cognitive ability connects with the remaining abilities in the network is allowed to vary both by the magnitude and the number of connections with the other cognitive abilities in the system. This implies that clustering can occur, that some variables may (therefore) share more common variance than others, and that networks can thus yield similar (even identical) model implied variance-covariance structures as factor models, albeit not necessarily so. Yet, the *explanation* of this hierarchical structure is thus radically different.

Within the field of intelligence, psychometric network modeling has initially been regarded as an exploratory means to help determine the number of factors in cognitive datasets (Golino & Demetriou, 2017). This is indeed a useful application. Yet, the potential of network analysis reaches further, because factor models can be regarded as network models in which the values of the pathways (edges) are constrained by the latent variables. In other words, factor models are nested within network models. The implementation of psychometric networks within the confirmatory framework may so be of help in distinguishing between factor analytical and network theories of intelligence, namely by comparing the fit of factor models to the fit of network models (and associated mutualism models). Other applications of confirmatory network modeling may lie in the comparison of networks among each other and the cross-validation of networks, e.g. ones that were established using exploratory network analysis in previous studies. This would parallel the transition from exploratory factor modeling to confirmatory factor modeling. The aim of the current paper is to evaluate such utility. To this end, we provide three applications of confirmatory network analysis using independent datasets, and which are all well-researched in the field of cognition and intelligence. An additional simulation study backs up the validity of the approach.

All confirmatory factor and network analyses were carried out in R (R Core Team, 2018) version 3.5.1 (Windows 10) using structural equation modeling package OpenMx version 2.11.5 (Boker et al., 2011) with Maximum Likelihood (ML) as estimator of the parameters. R codes are published on GitHub (https://github.com/kjkan/nwsem).

2. Application 1

An often-used approach to test competing factor analytical theories of intelligence against each other is to investigate published correlation matrices of psychometric intelligence batteries. The research into the factor structure of the Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV, Wechsler, 2008), to which some journals have devoted special issues (Tobin, 2013), constitutes a good example. On the basis of the correlations among the WAIS-IV subtests, some researchers have endorsed second order (*g*) factor models, albeit different ones (c.f., Wechsler, 2008; Benson, Hulac, & Kranzler, 2010; Canivez & Kush, 2013; Weiss, Keith, Zhu, & Chen, 2013), while others have rejected such models, in favor of an oblique first order (hence non-*g*) factor model (e.g., Ward, Bergman, & Hebert, 2012). Yet others have maintained that both should be rejected in favor of a bi-factor model (Gignac & Watkins, 2013). However, as mentioned, from a explanatory standpoint bifactor models of intelligence are considered unsatisfactory (Jensen, 1998; Hood, 2010).

In the light of competing *explanations* of the structure, the question remains if (theoretical satisfactory) factor models should be preferred over the network conceptualization of intelligence. To demonstrate that the implementation of psychometrics network can help answering this question, we extended the series of WAIS-IV analyses and compared the fits of (a) a measurement model, (b) a higher order *g* factor model, and (c) a non-saturated psychometric network model, with (d) a saturated model (interpreted as a network in which *all* variables connect directly to each other).

Table 1
Differences between traditional factor analysis and psychometric network analysis.

Difference	Traditional factor analysis	Psychometric network analysis
Formulation	(Hierarchical) factor structure	System of inter-relating variables
Terminology	Observed variables	Nodes (vertices)
	Correlations between observed variables	Edges
		A network is defined by a set of nodes connected by a set of edges of varying degrees of importance.
Preferred structure	A latent variable ‘general factor’ is at the apex.	By partial correlations.
	More specific factors are arranged below.	Two subtests are related, and that relation cannot be due to any of the other variables in the network.
Subtest level interpretation	Each subtest has one or more standardized ‘factor loadings’ (regression coefficient on the factors in the model) ranging from −1 to 1. These loadings determine the extent to which a factor influences individual differences in subtest scores.	Each node or vertice (observed variable) is indexed by ‘centrality’. Centrality indices identify ‘the most fundamental variables’ in the network
Subtest cluster interpretation	The labels of the latent factors are derived from the subtest labels and the way the observed variables load on the factors.	Labels of observed clusters (‘communities’) are derived from the way the observed variables group in the network.
Concrete example	Vocabulary and performance on mathematical tasks are correlated due to their common dependency on the variable g, which has not been observed, but is hypothesized to exist.	Any partial correlation between vocabulary and performance on mathematical tasks denotes a direct dependency, i.e. after taking into account possible shared dependencies on the other variables in the network.

2.1. Method

2.1.1. Participant sample

The sample consisted of 1800 individuals (aged 16 to 89 years old) who participated in the standardization of the WAIS–IV and who completed all 15 subtests of the battery (Wechsler, 2008).

2.1.2. Measures

The subtests were: (1) *Similarities* (SI), (2) *Vocabulary* (VC), (3) *Information* (IN), (4) *Comprehension* (CO), (5) *Block Design* (BD), (6) *Matrix Reasoning* (MR), (7) *Visual Puzzles* (VP), (8) *Figure Weights* (FW), (9) *Picture Completion* (PC), (10) *Digit Span* (DS), (11) *Arithmetic* (AR), (12) *Letter-Number Sequencing* (LN), (13) *Symbol Search* (SS), (14) *Coding* (CD), and (15) *Cancellation* (CA). For detailed information concerning the psychometric properties of the WAIS-IV, we refer to its Technical and Interpretative Manual, which also contains the correlation matrix among the subtest scores (Wechsler, 2008, Table 5.1, p.62).

2.1.3. Statistical approach

2.1.3.1. Factor analysis. The following confirmatory factor models were fitted: (1) a measurement model, i.e. an oblique first order-four-factor model which includes *Verbal Comprehension* (indicated by SI, VC, IN, and CO), *Perceptual Reasoning* (indicated by BD, MR, VP, FW, and PC), *Working Memory* (indicated by DS, AR, and LN), and *Processing Speed* (indicated by SS, CD, and CA) as the four correlated latent variables; (2) a second order hierarchical factor model that explained the correlation between these latent variables by postulating a shared dependency of these variables on a second order latent variable, '*g*'.

2.1.3.2. Psychometric network analysis. An exploratory psychometric network was established following the guidelines and recommendations (Costantini et al., 2015; Epskamp, Borsboom, & Fried, 2018) that are used widely in network psychometrics. This involved the computation of a Gaussian Graphical Model (GGM) of the full partial correlation matrix (the use of full partial correlations rather than Pearson correlations implies that any connection between two variables cannot be attributed to any of the other variables in the model). This can be considered the saturated network model. Subsequently, by means of qgraph's built-in unregularized GGM model search procedure, the network model was adjusted for the presence of false positive edges, i.e., partial correlations between pairs of observed variables that probably do not differ from 0. The result was a more sparse (hence nonsaturated) network model.

2.1.3.3. Networks as structural models. Package qgraph is under development and does not yet provide many of the commonly used fit indices. However, because the matrix algebraic properties of psychometric networks have been explicated (Epskamp, Rhemtulla, & Borsboom, 2017; see also Appendix A), it is possible to obtain these indices by implementing the networks in OpenMx (Boker et al., 2011), which is a structural equation R package that uses matrix algebraic expressions.

Since a full network model is a saturated model and will thus yield perfect fit, this type of network modeling might thus not be the most interesting from a theoretical perspective. Yet those models can be considered as the null-models against which factor models are tested. As structural equation software allows for (equality and inequality) constraints, specific parameters can be set to 0 (either a priori or post hoc). In OpenMx this utility of SEM can be applied to elements in the partial correlation matrix, so that the fit of sparser, hence theoretically more interesting networks can be assessed. With the purpose to demonstrate this utility, we implemented in OpenMx the network that resulted from the exploratory analysis above. Ideally, in a true confirmatory analysis, one would have specified the network a priori (see Applications 2 and 3).

2.1.3.4. Model selection by model comparison. Model selection involves multiple steps: (1) Model *testing*, (2) model *fitting*, and (3) relative fit *comparison*, and in that order (Kline, 2015). The first step usually consists of a (central or noncentral) χ^2 test that is applied to all the models under consideration. As a statistical test, the χ^2 assesses whether there is a significant discrepancy between the observed and model implied variance-covariance matrices, where significant χ^2 values represent a significant discrepancy. The χ^2 statistic can also be used as a measure of fit (in step 2): A *nonsignificant* result implies a good model fit, i.e. a satisfactory reduction of the data, because the model could not be rejected as constituting the data generating mechanism. As a measure of fit, the outcome of the χ^2 test is sensitive to sample size, so that a *significant* result is not necessarily indicative of a 'bad' fit; the model may still be considered to summarize the data well enough. Because of this sensitivity it is common practice in factor analysis and structural equation modeling to judge models (also) according to alternative fit indices.

Per the recommendations in OpenMx (Boker et al., 2011), we obtained in our analyses: (1) The Non-normed Fit Index (NNFI), which represents the proportion of total covariance among the observed variables that is explained by a target model with the null model as a baseline model; (2) the Comparative Fit Index (CFI), which is based on the non-central χ^2 distribution, and accounts for sample size; and (3) the Root Mean Square Error of Approximation (RMSEA) together with its associated 95% Confidence Intervals (CI) and *P*-values. The RMSEA estimates the extent to which the model, with unknown, but optimally chosen parameter values would fit the population covariance matrix if it were available. The following (standard) cut-off values were applied in order to conclude a model fits the data well: NNFI > 0.90 and CFI > 0.93. RMSEA values of < 0.01, < 0.05, and < 0.08 were considered to reflect excellent, good, and acceptable fit, respectively (Byrne, 1994; Hu & Bentler, 1999; MacCallum, Browne, & Sugawara, 1996).

To assess relative fit (step 3), the log likelihood is obtained, which is a maximized likelihood function expressed negatively, such that nested models can be compared by means of log-likelihood ratio testing (Wilks, 1938). Like the χ^2 test in step 1, this log-likelihood ratio test is also sensitive to sample size, however. For non-nested models a true statistical test is not available, but these (and nested models as well) can be compared using fit criteria such as Akaike's information criterion (AIC; Burnham & Anderson, 2004), the Bayesian information criterion (BIC; Schwarz, 1978), or adjusted versions of the latter, e.g. the sample adjusted BIC (SABIC; Sclove, 1987). These fit criteria are all functions of both the log-likelihood and the number of freely estimated parameters in the model. The AIC estimates the relative quality (information loss; where lower values represent a better fit). The two Bayesian criteria extend the AIC by penalizing the number of parameters more strongly.

2.2. Results

The model fitting results are provided in Table 2. This table highlights that at the nominal significance level of $\alpha = 0.05$, the χ^2 statistic would reject all models under consideration, but also that according to the alternative fit measures, the fits of the measurement model and the higher order *g* factor model were acceptable, and the fits of the network model good. In line with those results, the relative fit information criteria (AIC, BIC, SABIC) indicated that the sparse network model yielded better relative fits than the factor models, hence even after penalizing for model complexity.

On the basis of those results, we would come to the conclusion that the *g* model deserved rejection in favor of the network model. A risk, however, is that we have arrived at an overparameterized solution.

2.3. Validation by simulations

To investigate if psychometric network modeling (through R package qgraph) yielded an overparameterized solution, and showed

Table 2
Showing the fit statistics of obtained in Application 1 (analysis of the Wechsler Adult Intelligence Scale – Fourth Edition).

Model	-2LL	χ^2	df	*P*	AIC	BIC	SABIC	CFI	NNFI	RMSEA	$CI95_l$	$CI95_u$	P_{RMSEA}
Saturated model	11,568.16	0	0	1.00	240.00	899.47	518.23	1.00	1.00	0.00	–	–	–
Measurement model	12,464.85	896.70	99	< 0.01	968.70	1.166.54	1.052.17	0.95	0.93	0.07	0.068	0.079	0.00
Second order *g*-model	12,490.71	922.56	101	< 0.01	990.56	1.177.41	1.069.39	0.95	0.93	0.07	0.068	0.079	0.00
Network	**11,698.10**	**129.94**	**71**	**< 0.01**	**257.94**	**609.65**	**406.33**	**1.00**	**0.99**	**0.02**	**0.014**	**0.028**	**1.00**

Note. Abbreviations: -2LL: minus 2 times the log-likelihood, df: degrees of freedom, $CI95_l$ and CI95u: lower and upper boundaries of the 95% confidence interval of the RMSEA value; P_{RMSEA}: the *P*-value associated with this interval. Preferred model bold faced and underlined.

the better fit for that reason, we conducted an additional Monte Carlo simulation study.

2.3.1. Method

Multivariate normal sample data were generated according to the observed WAIS-IV correlation matrix (saturated model) as well as according to the factor and network models described above. Next, all these models were fitted on all generated sample data sets, after which we determined in each sample which model was selected by the AIC, BIC and SABIC as the best model. So, for instance, in one series of simulations we generated sample data according to the network model above and fitted in each of these samples (1) this network model, (2) the measurement model, and (3) the higher order *g* factor model. In a parallel analysis, we did not fit the original network, but derived - for each sample separately - a new network through qgraph and fitted that network on the data. This was repeated for situations in which the higher order *g* model, the measurement model and the saturated model were the true models.

A single series consisted of 1000 runs. The sample size in each run equaled the original WAIS-IV sample size (n = 1800, see above). The sample size and model implied variance-correlation matrices were used as input to the multivariate normal simulation function 'mvnorm' in R package MASS (Ripley, 2002). Means remained unmodeled and were set at 0 (in the population).

Apart from the relative fit criteria, we obtained the outcomes of the χ^2 test and the alternative fit indices described above. The χ^2 test was expected to yield approximate uniform distributions for the true model, so that at a nominal significance level of $\alpha = 0.05$, the true model would be rejected in about 5% of the cases (the Type I error frequency). The χ^2 test was furthermore expected to not always reject competing models (leading to Type II errors). If overparameterization would occur, the AIC, BIC, and SABIC were expected to favor the true model in general (but perhaps not in every single case).

2.3.2. Results

The χ^2 test and fit criteria performed as expected. When the true model was fitted, *P*-values pertaining to the χ^2 test statistic of the true model were indeed approximately uniformly distributed. If a wrong model was fitted, *P*-values were generally skewed to the right. An exception was when the true model was the higher order *g* model while the measurement model was fitted to the data. In that case, *P*-values were also approximately uniformly distributed (in line with the nestedness).

The results of model selection procedure are presented in Table 3. They highlight that the fit criteria performed well and picked out the true model - i.e. the higher order *g* model, measurement model or network - in the large majority of cases. So, for example, when the true model was a higher order factor model and the measurement model and network model would thus also provide a solution, the latter two were generally (correctly) rejected, because of their unnecessary complexity.

When the unmodeled structure (saturated model) was the input, the AIC identified the saturated model as the true data generating mechanism, but the BIC and SABIC, which punish stronger for complexity, then suggested the nonsaturated network to be the model of preference. This signified that the network model (df = 71) implied matrix and the observed WAIS matrix were close. Since the network model clearly replicated over each independent sample of the population, that close fit was unlikely due to overparameterization.

2.4. Conclusion

With respect to the description of the WAIS IV variance-covariance structure, we can conclude that the network model provided a better *description* than the *g* model. From a substantive perspective, we may conclude that the result provides evidence against *g* in favor of a network conceptualization of intelligence. This in turn fits with the mutualism theory of intelligence. Hence between the two *explanations* - the *g* hypothesis and the mutualism hypothesis - mutualism is the preferred one.

3. Application 2

Application 1 demonstrated the key advantage of implementing a network as a structural equation model, namely the possibility to obtain commonly used fit statistics. Application 2 aims to demonstrate a second advantage, namely the opportunity to cross-validate networks, extending psychometric network analysis with truly confirmatory techniques. Similar to investigations in which a previously established factor model is tested in other (sub)samples, networks established in the one (sub)sample - e.g. by exploratory network analysis in qgraph - can now be subjected to statistical testing in other (sub)samples.

3.1. Method

3.1.1. Participant sample

The total sample consisted of 301 seventh and eighth grade children who participated in a study conducted by Holzinger and Swineford (1939).

3.1.2. Measures

The original, full dataset contained scores on 26 tests, which aimed to indicate the children's *Verbal ability*, *Mathematical-ability*, *Spatial ability*, *Memory capacity*, and *Mental speed*. In the literature, a smaller subset of 9 variables is used more widely, e.g. in SEM-teaching material. This subset is freely available in multiple SEM programs, including LISREL (Jöreskog & Sörbom, 1989), Mplus (Muthén & Muthén, 2012), and AMOS, (Arbuckle, 2014) as well as in R SEM packages OpenMx (Boker et al., 2011) and lavaan (Rosseel, 2012). The subtests that comprise this subset include *Visual perception* (VIS), *Cubes* (CUB), *Lozenges* (LOZ), *Paragraph comprehension* (PCM), *Sentence completion* (SCM), *Word meaning* (WM), *Speeded addition* (SA), *Speeded counting of dots* (SCD), and *Speeded discrimination straight and curved capitals* (SDC).

3.1.3. Statistical approach

In short, the sample was split randomly in two subsamples (using R function 'sample'). This was possible since the Holzinger and Swineford data concerned raw scores rather than a correlation or covariance matrix. Next, on the basis of the covariance matrix in Subsample 1 an exploratory psychometric network model was extracted using the same procedure as in Application 1. Subsequently, this network was fitted on

Table 3
Showing the results of the simulation study, i.e. the frequencies the true model was chosen according to the AIC, BIC, and SABIC.

		Preferred model				Preferred model			
		Higher order	Measurement	True network	Saturated	Higher order	Measurement	Extracted network	Saturated
True model	Criterion								
Higher order	AIC	86.7%	13.3 %	0%	0%	86.7%	13.3 %	0%	0%
	BIC	99.8%	0.2%	0%	0%	99.8%	0.2%	0%	0%
	SABIC	98.2%	1.8%	0%	0%	98.2%	1.8%	0%	0%
Measurement	AIC	0.1%	99.9%	0%	0%	0.1%	99.9%	0%	0%
	BIC	8.2%	91.8%	0%	0%	8.2%	91.8%	0%	0%
	SABIC	0.9%	99.1%	0%	0%	0.9%	99.1%	0%	0%
Network	AIC	0%	0%	100%	0%	0%	0%	100%	0%
	BIC	0%	0%	100%	0%	0%	0%	100%	0%
	SABIC	0%	0%	100%	0%	0%	0%	100%	0%
Saturated	AIC	0%	0%	0%	100%	0%	0%	38.1%	61.9%
	BIC	0%	0%	100%	0%	0%	0%	100%	0%
	SABIC	0%	0%	98.6%	1.4%	0%	0%	100%	0%

the (raw) data in Subsample 2.

In addition to the network model, and in line with the procedure followed in Application 1, we also fitted a measurement model and higher order *g* factor model in the two subsamples. The measurement model included the three (correlated) first order factors *Verbal Ability* (indicated by the three verbal tests PCM, SCM, and WM), *Spatial Ability* (indicated by the three spatial tests VIS, CUB, and LOZ), and *Speeded addition* (indicated by the three speed tests SA, SCD, and SDC). In the second order factor model the correlations between the three first order factors were explained by a second order general factor '*g*'. The models were judged on the basis of the same fit criteria as in Application 1. We note that the measurement model and *g* factor model were expected to yield the same results as the measurement model, as the number of first order factors was only three.

3.2. Results

The results of Application 2 are summarized in Table 4, from which it can be obtained that they corroborated the results from Application 1: In subsample 1 the model fits of the measurement model and the *g* model were acceptable and the network model fitted well. In subsample 2; the network model fitted not as well as in subsample 1, but the fits were still acceptable and according to the relative fit indices the network model still outperformed the *g* model.

3.3. Conclusion

As illustrated, confirmatory network modeling provides a means to cross-validate networks. In the example above, a network established in the one subsample fitted in the next subsample, meaning that the network replicated over the subsamples; the skeleton of the network could thus be assumed invariant over the subsamples. Not only in the subsample from which the network was extracted, but also in the subsample in which the network was cross-validated, the network outperformed the factor models.

Although the skeleton replicated, parameter values may have not. In Application 3, we illustrate how to test if (two or more) networks can be also be considered invariant in their *parameter values*. This utility is first of all of interest for psychometric network modelers, but can also be conceived of providing additional tests within research into measurement invariance.

4. Application 3

Apart from illustrating the possibility to test for parameter invariance, we also aimed to address a more substantive question, namely if there are any changes in the (factor or network) structure during cognitive aging.

Table 4
Showing the fit statistics of obtained in Application 1 (analysis of the Wechsler Adult Intelligence Scale Fourth Edition).

Fit in Subsample 1

Model	-2LL	χ^2	df	*P*	AIC	BIC	SABIC	CFI	NNFI	RMSEA	$CI95_l$	$CI95_u$	P_{RMSEA}
Saturated model	3722.31	0.00	0	1.00	3830.31	3992.89	3821.99	1.00	1.00	0.00	–	–	–
Measurement model	3761.68	39.36	24	0.02	3821.68	3912.00	3817.05	0.96	0.94	0.07	0	0.107	0.229
Higher order *g* model	3761.68	39.36	24	0.02	3821.68	3912.00	3817.05	0.96	0.94	0.07	0	0.107	0.229
Network model	3742.77	20.46	20	0.43	3810.78	3913.13	3805.53	1.00	1.00	0.01	0	0.079	0.794

Refit in Subsample 2

Model	-2LL	χ^2	df	*P*	AIC	BIC	SABIC	CFI	NNFI	RMSEA	$CI95_l$	$CI95_u$	P_{RMSEA}
Saturated model	3609.51	0.00	0	1.00	3717.51	3880.44	3709.54	1.00	1.00	0.00	–	–	–
Measurement model	3684.44	74.93	24	0.00	3744.44	3834.96	3740.01	0.90	0.84	0.12	0.083	0.155	0.000
Higher order *g* model	3684.44	74.93	24	0.00	3744.44	3834.96	3740.01	0.90	0.84	0.12	0.083	0.155	0.000
Network model	3646.76	37.25	20	0.01	3714.76	3817.35	3709.75	0.96	0.94	0.08	0.025	0.119	0.127

Note. Abbreviations: -2LL: minus 2 times the log-likelihood, df: degrees of freedom, $CI95_l$ and $CI95_u$: lower and upper boundaries of the 95% confidence interval of the RMSEA value; P_{RMSEA}: the *P*-value associated with this interval.

4.1. Method

4.1.1. Participant sample

The sample consisted of individuals who participated in the Midlife in the United States (MIDUS) national study (Lachman, Agrigoroaei, Tun, & Weaver, 2014). At the start of this large-scale study the number of participants (then aged 24 to 75) equaled 7100 (Brim, Ryff, & Kessler, 2004). The cognitive performance measures were collected 9 years later (together with a rich assortment of other variables, see Lachman & Tun, 2008; Tun & Lachman, 2006; Tun & Lachman, 2008). These cognitive scores pertain to 3779 individuals.

To test for possible age related changes, the sample was divided into age groups, based on established age cut-offs used in the literature (Crowley et al., 2016). The group *younger* consisted of participants of age 35 to 54 ($N = 1876$), the group *middle* age 55 to 64 ($N = 1000$) and *older* 65 or older (up to 86) ($N = 903$).

4.1.2. Measures

Measures comprised the scores on the seven subtests of the Brief Test of Adult Cognition by Telephone (BTACT; (Lachman et al., 2014). These subtests assess (I) *immediate recall*; (II) *delayed recall*; (III) *working memory span* (backward digit span); (IV) *verbal fluency* (category fluency); (V) *inductive reasoning* (completing a pattern in a series of 5 numbers); (VI) *processing speed* (the number of digits produced by counting backward from 100 in 30 s); and (VII) *attention switching and inhibitory control* (by the Stop and Go Switch Task; Tun & Lachman, 2008).

For detailed information concerning the psychometric properties of the BCTAT, we refer to Lachman et al. (2014). In short, the reliability of the BCTAT has been established by means of parallel-forms and test-retesting; the concurrent validity has been based on correlations with a full-test face-to-face assessment and the external validity on the correlations between BCTAT scores and demographic factors and indices of health; the construct validity has been investigated by confirmatory factor modeling. This factor modeling revealed two correlated factors ($r = 0.34$), which are commonly interpreted as *Episodic memory* (indicated by immediate and delayed word recall) and *Executive functioning* (indicated by the remaining subtests).

4.1.3. Statistical approach

The statistical approach is detailed below, but can be summarized as follows: For each age group, we first derived descriptive statistics, including means, variances, and correlation matrices. As these have not been published previously they are presented in Table 5. Further analysis involved again the fit of factor models and psychometric network models. Next, these models were implemented in a multigroup fashion. Provided models were tenable in each age group, we tested for parameter invariance over the age groups by introducing equality constraints. To assess fit, the same indices were used as in Applications 1 and 2.

4.1.3.1. Factor analysis. Following Lachman et al. (2014), we fitted in each subsample a measurement model, i.e. an oblique first order-two-factor model with the factors episodic memory and executive functioning as two correlated latent variables. A second order factor model was not fitted since the number measurement model was only two, leaving this model unidentified. In addition, if we would identify it by constraining the second order loadings to be equal, this would yield the exact same fit as the fit of the measurement model.

4.1.3.2. Network analysis. In the group younger we extracted a network model using the same guidelines as in Application 1 and 2. Subsequently, this network was fitted in the groups *middle* and *older*, after which we obtained the fit statistics of these unigroup models.

4.1.3.3. Model comparison. Next, the aim was to specify the factor model and the network model in a multigroup fashion, after which we tested if parameters (factor loadings in the factor model, edges in the network model) could be assumed equal over the age groups.

4.2. Results

4.2.1. Descriptive statistics

The visualization of the full partial correlation networks (Fig. 2) showed that Lachman et al., 2014 ‘s distinction between episodic memory and executive functioning appeared meaningful, since immediate recall and delayed recall clustered more strongly together than with the other measures.

4.2.2. Model fitting results

As the modeling statistics in Table 6 show, the fit of the measurement model was acceptable in the group *younger*, inconclusive in the group *middle*, and not acceptable in the group *older*. This result made that testing the assumption of equal factor loadings would not have any meaning. The fits of the network model established in the group *younger* were good in all three groups. The information criteria (AIC, BIC, SABIC) corroborated these findings. They revealed better relative fit for the network model, not only in group *younger* from which the network was extracted, but also in the groups *middle* and *older*.

According to the multigroup modeling results edges could not be constrained without a significant reduction in fit: $\Delta\chi^2$ (28) = 60.60, $P < 0.001$. So although configural network invariance was established, the stricter assumption of parameter invariance needed rejection.

From a substantive perspective, we conclude that over age (or cohorts) a single network skeleton describes the structure among the variables best. Yet, there are age changes (or cohort differences) in at least one of the edges.

5. Discussion

We fitted multiple confirmatory factor analytic and psychometric network models on three independent, well-researched datasets. Our results firstly highlight the utility of extending the kind of psychometric network analysis (Borsboom, 2008; Epskamp et al., 2012) that rapidly gained popularity in the fields of psychopathology (Fried et al., 2017) and entered recently the field of cognition and intelligence (Golino & Demetriou, 2017; van der Maas et al., 2017). As demonstrated, the implementation of network models within a confirmatory (structural equation) modeling framework (Boker et al., 2011; Epskamp et al., 2017) permits, for instance, (1) the comparisons among factor and networks models, which can assist in the comparison of a priori theoretically driven models, (2) the comparison of networks over groups, and (3) the combination of these two.

From a descriptive viewpoint concerning individual differences in cognitive performance, the major finding of interest was that the psychometric networks provided better descriptions of the data than previously established confirmatory factor analytic models. Additional simulations showed this is unlikely due to overparameterization. In view of substantive theory, our results imply that the hypothesis of an underlying general factor of intelligence is not required in order to explain the pattern of correlations between the different cognitive performance measures. More strongly, the current results provide an empirical argument against *g* theory (e.g. Jensen, 1998) favoring the mutualism theory of intelligence (van der Maas et al., 2006). The latter posits that positive associations between cognitive abilities arise through reciprocal dynamical interaction between those abilities during development, and that this is a sufficient explanation.

According to mutualism theory, interactions between any two cognitive abilities can be direct (either one directional or bidirectional) and/or indirect (directional or bidirectional). Some of the interactions may even be zero or negative; as long as the majority of the interactions is positive, a positive manifold of Pearson correlations can be expected (van der Maas et al., 2006). The presence of variety in edge strength

Table 5
Partial and Spearman's correlation matrices among the Brief Test of Adult Cognition by Telephone (BTACT) tests.

	Immediate Recall	Delayed Recall	Digit Span	Category Fluency	Number series	Backward counting	SGST
Group Younger (age 35–54; n = 1876)							
Immediate recall		0.73	0.10	0.07	0.01	0.03	0.01
Delayed recall	0.76		0.07	0.01	0.05	−0.04	0.04
Digit span	0.29	0.28		−0.01	0.19	0.16	0.01
Category fluency	0.19	0.17	0.15		0.20	0.14	0.10
Number series	0.20	0.20	0.31	0.32		0.25	0.07
Backward counting	0.16	0.13	0.29	0.29	0.40		0.31
SGST	0.14	0.14	0.16	0.23	0.25	0.40	
Group Middle (age 55–64; *n* = 1000)							
Immediate		0.74	0.09	0.05	0.05	0.07	0.01
Delayed	0.76		0.08	0.02	0.01	−0.05	−0.02
Digit span	0.30	0.27		0.01	0.20	0.10	0.02
Category fluency	0.19	0.15	0.19		0.20	0.22	0.11
Number series	0.23	0.18	0.33	0.37		0.26	0.04
Backward counting	0.19	0.12	0.27	0.41	0.44		0.38
SGST	0.10	0.05	0.15	0.29	0.26	0.48	
Group Older (age 65 and older; n = 903)							
Immediate		0.73	0.20	0.07	0.05	−0.05	0.11
Delayed	0.78		0.08	0.05	−0.01	0.06	−0.09
Digit span	0.43	0.38		−0.02	0.18	0.06	0.03
Category fluency	0.24	0.22	0.16		0.16	0.13	0.16
Number series	0.24	0.20	0.31	0.30		0.34	−0.05
Backward counting	0.19	0.18	0.23	0.33	0.44		0.39
SGST	0.18	0.11	0.16	0.30	0.21	0.46	

Note. To the right, upper triangle consists of partial correlations and lower triangle of Pearson's r correlation coefficients. Statistically significant (P < 0.01) values underlined for clarity.

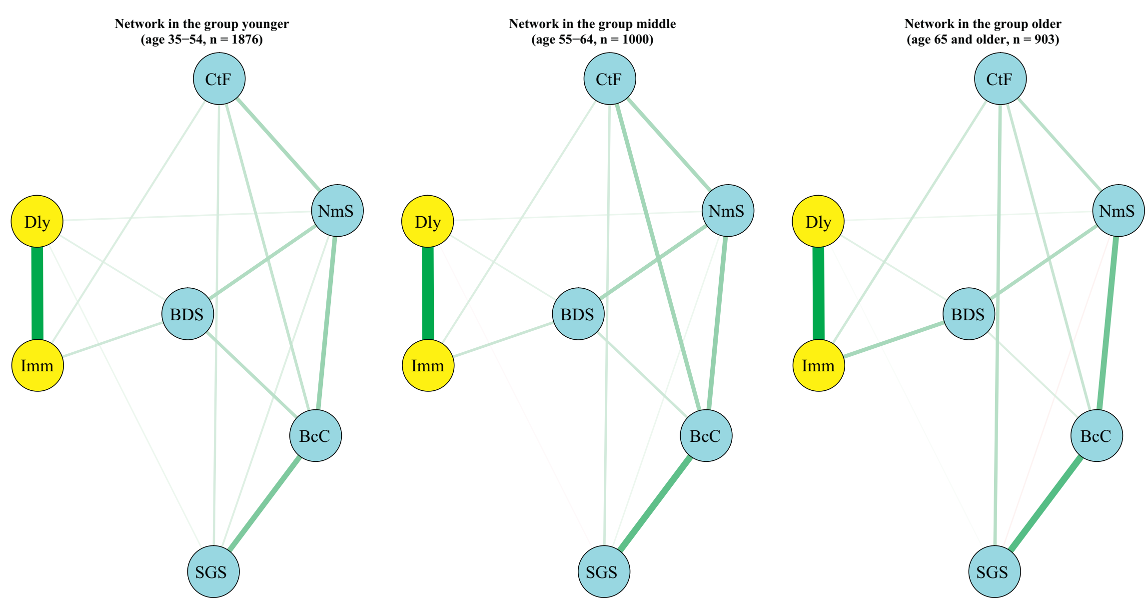


Fig. 2. showing networks by age group using the coordinates of the age group invariant restricted network. Edges were estimated freely within the age groups. The skeleton is identical over the groups, but parameter values could not be assumed all equal over the groups, so at least 1 pair differs. Note. The edges represent partial correlations. The thicker the edges, the stronger the partial correlation. Abbreviations: Imm: Immediate recall, Dly: Delayed recall, BDS: Backward Digit Span, CtF: Category Fluency, NmS: Number Series, BcC: Backward Counting, SGS: Performance on the Stop and Go Switch Task.

and of sparsity is particularly interesting, as it links to observed hierarchical organization of intelligence itself, but also to the hierarchical organization among (brain) variables that related to intelligence (e.g. Clauset, Moore, & Newman, 2008; Taylor, Hobbs, Burroni, & Siegelmann, 2015).

We note that we do not argue against a hierarchical conceptualization of intelligence from a *descriptive* (or constructivist) perspective. Rather the current results are more consistent with the notion that such hierarchical organization is the result of dynamical interaction between the subsystems that underlie cognitive performance. In our view, the hierarchy is an emergent property of the system as a whole. Of importance in future research is first to identify 'mediational nodes', and, next, to specify how the effects of (non-psychometric) variables are mediated by those nodes. Variables of interest

Table 6
Showing the fit statistics of the statistical modeling in Application 3.

Uni-group model (group)	-2LL	χ^2	df	P	AIC	BIC	SABIC	CFI	NNFI	RMSEA	$CI95_l$	$CI95_u$	P_{RMSEA}
Saturated (younger)	52,772.85	0.00	0	1.00	52,842.8	53,036.6	52,925.5	1.00	1.00	0.00	–	–	–
Measurement (younger)	52,927.58	154.73	13	< 0.001	52,971.6	53,093.4	53,023.5	0.95	0.93	0.08	0.064	0.089	< .0.01
Network (younger)	52,777.36	4.51	7	0.72	52,833.4	52,988.4	52,899.4	1.00	1.00	< 0.001	< 0.001	0.025	> 0.99
Saturated (middle)	28,050.67	0.00	0	1.00	28,120.7	28,292.4	28,181.3	1.00	1.00	0.00	–	–	–
Measurement (middle)	28,163.33	197.13	13	< 0.001	28,207.3	28,315.3	28,245.4	0.95	0.91	0.09	0.070	0.106	< 0.001
Network (middle)	28,064.16	13.49	7	0.06	28,120.2	28,257.6	28,168.6	1.00	0.99	0.03	< 0.001	0.059	0.90
Saturated (older)	25,024.72	0.00	0	1.00	25,094.7	25,262.9	25,151.8	1.00	1.00	0.00	–	–	–
Measurement (older)	25,221.85	112.66	13	< 0.001	25,265.8	25,371.6	25,301.7	0.89	0.83	0.13	0.107	0.144	< 0.001
Network (older)	25,043.28	18.56	7	0.01	25,099.3	25,233.8	25,144.9	0.99	0.98	0.04	0.014	0.071	0.66

Multi-group model *(nested model)*	-2LL	χ^2	df	P	AIC	BIC	SABIC	CFI	NNFI	RMSEA	$CI95_l$	$CI95_u$	P_{RMSEA}
Saturated	105,848.2	0.00	0	1.00	106,058.2	106,713.1	106,379.5	1.00	1.00	0.00	–	–	–
Network	**105,884.8**	**36.56**	**21**	**0.019**	**106,052.8**	**106,576.7**	**106,309.8**	**1.00**	**0.99**	**0.01**	**0.003**	**0.023**	**> 0.99**
(*Invariant edges*)	105,945.4	97.15	49	< 0.001	106,057.4	106,406.7	106,228.7	0.99	0.99	0.02	0.010	0.022	> 0.99

Note. Abbreviations: -2LL: minus 2 times the log-likelihood, df: degrees of freedom, $CI95_l$ and $CI95_u$: lower and upper boundaries of the 95% confidence interval of the RMSEA value; P_{RMSEA}: the P-value associated with this interval. Preferred model bold faced and underlined.

may include, genetic variants and brain correlates, for instance, but also societal factors, such as education. The combination of non-psychometric and psychometric variables may so shed further light on the etiology of the hierarchical organization.

Apart from the fact that psychometric network models outperformed traditional factor models, we obtained additional findings of theoretical interest. The variance-covariance structure of the MIDUS BTACT could not be assumed age invariant. This suggests some changes in the variance covariance structure, hence in factor structure across age (or cohorts). As the network's skeleton could be assumed invariant, the shift in the values of edges may denote a shift in interaction strengths across age (or generations), which then may be result of natural maturation (van der Maas et al., 2006) or of changes in environmental requirements (Dickens & Flynn, 2001; van der Maas et al., 2017), for instance.

5.1. Limitations

A limitation of the current analyses was that all data concerned cross-sectional data, precluding the study of intra-individual differences. To distinguish optimally between g theory and the mutualism theory, we therefore advocate the type of longitudinal modeling used by Hofman et al. (2018) and Kievit et al. (2017). Of additional advantage is that in these models asymmetries in mutualistic interactions can be studied. Future research along these lines can be further extended by including broader sets of cognitive variables (rather than sets of two).

It should also be noted that although we applied network modeling using *confirmatory* techniques, (different, competing) a priori theoretical network models were not tested. Application 1 in particular must still be considered as being *explorative* in nature. In the literature, this type of structural equation modeling – a utilization that nevertheless contributes greatly to the clarification and the development of theories – is referred to as 'the exploratory mode' of confirmatory modeling (Raykov & Marcoulides, 2012). Results always requires replication across other samples from the same population, so that the overall results can be considered trustworthy. The reason is that results obtained from one study are limited generalizable; chance factors may have led to a particular dataset. Our simulation study showed that the risk of overparameterization can nevertheless be considered low. Possibly the risk increases in relatively small samples, so we advance the use of relatively large samples.

5.2. General conclusion

Overall, the current study promotes confirmatory psychometric network analysis, in the field of cognition and intelligence in particular, and in differential psychology in general. Future research along these lines is warranted in order to distinguish between factor analytical and network interpretations on the etiology of intra- and interindividual differences.

With respect to the debate concerning the theoretical status of g, we conclude the following. We do not exclude the presence of common or general influences, e.g. of certain genetic variants or environmental variables like exposure to education. The question to be answered is more *how* such effects could have arisen: Are they the result of dynamical reciprocal interactions or are they due to a single mediating variable g which has never been found to exist? The evidence from the current series of studies argues clearly against the latter and therefore against mainstream g theory. They favor the mutualism theory of intelligence.

Declaration of interest

Over 5 years ago Professor Stephen Z Levine received research support,and/or consultancy fees and/or travel support from Maccabi HealthcareServices, F. Hoffmann-La Roche, Shire Pharmaceuticals and Eli Lillythat are not relevant to this study.

Appendix A. Appendix

In general, a confirmatory (n-) factor model aims to describe the observed variance-covariance matrix (among m variables), $\mathbf{\Sigma}$, as $\hat{\mathbf{\Sigma}} = \mathbf{\Lambda\Phi\Lambda}^{\mathrm{T}} + \mathbf{\Theta}$ (LISREL notation; Jöreskog & Sörbom, 1989), where $\mathbf{\Lambda}$ is the $(m \times n)$ matrix containing the factor loadings, $\mathbf{\Phi}$ the $(n \times n)$ variance-covariance matrix of the n latent factors, and $\mathbf{\Theta}$ the $(m \times m)$ variance-covariance matrix of the residuals (of the observed variables).

Psychometric network models aim to describe $\mathbf{\Sigma}$ as $\hat{\mathbf{\Sigma}} = \mathbf{\Delta}(\mathbf{I} - \mathbf{\Omega})^{-1}\mathbf{\Delta}$ (Epskamp, Rhemtulla, & Borsboom, 2016), where $\mathbf{I}$ is an $(m \times m)$ identity matrix, $\mathbf{\Omega}$ is an $(m \times m)$ matrix containing partial correlations (but with 0's on the diagonal, such that $\mathbf{I} + \mathbf{\Omega}$ is the actual partial correlation matrix). Matrix $\mathbf{\Delta}$ is a $(m \times m)$ diagonal matrix containing scaling parameters.

The mutualism model aims to describe $\mathbf{\Sigma}$ as $\hat{\mathbf{\Sigma}} = (\mathbf{I} - \mathbf{M})^{-1}\mathbf{\Psi}(\mathbf{I} - \mathbf{M})^{-\mathrm{T}}$ (van der Maas et al., 2006), where $\mathbf{I}$ is again an $(m \times m)$ identity matrix and $\mathbf{M}$ the matrix containing the mutualistic weights, m_{ij}, i.e. regression weight of variable i on variable j. In the standard mutualism model matrix $\mathbf{\Psi}$ is a diagonal matrix containing the variances of limiting capacities. The formula of the covariance structure predicted by the mutualism model is essentially the same as the general formula for path models (Wright, 1934), in which $\mathbf{\Psi}$ is not (necessarily) diagonal.

More generally models can be framed within the Reticular Action Model (RAM; McArdle & McDonald, 1984): $\hat{\Sigma} = \mathbf{F}(\mathbf{I} - \mathbf{A})^{-1}\mathbf{S}(\mathbf{I} - \mathbf{A})^{-\mathrm{T}}\mathbf{F}^{\mathrm{T}}$. Here **F** is an $(m \times (m + n))$ filter matrix (consisting of a $(m \times m)$ identity matrix augmented with a $(m \times n)$ matrix containing zeroes); **A** an (asymmetric) matrix containing all directed pathways between all (observed and latent) variables; **S** a symmetric matrix containing all undirected pathways between all (observed and latent) variables. In network models, for instance, **S** is thus further specified as $\mathbf{\Delta}(\mathbf{I} - \mathbf{\Omega})^{-1}\mathbf{\Delta}$, whereas Matrix **F** is restricted to be an identity matrix and **A** a matrix containing zeroes.

References

Arbuckle, J. L. (2014). *Amos (Version 23.0) [Computer Program].* Chicago: IBM SPSS.

Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for thomson's bonds model of intelligence. *Psychological Review, 116*(3), 567.

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the wechsler adult intelligence Scale—Fourth edition (WAIS-IV): What does the WAIS-IV measure? *Psychological Assessment, 22*(1), 121.

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., & Bates, T. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika, 76*(2), 306–317.

van Borkulo, C., Boschloo, L., Borsboom, D., Penninx, B. W., Waldorp, L. J., & Schoevers, R. A. (2015). Association of symptom network structure with the course of depression. *JAMA Psychiatry, 72*(12), 1219–1226.

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology, 64*(9), 1089–1108.

Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology, 9*, 91–121.

Brim, O. G., Ryff, C. D., & Kessler, R. C. (2004). *The MIDUS national survey: An overview.* University of Chicago Press.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261–304.

Burt, C. (1909). Experimental tests of general intelligence. *British Journal of Psychology, 1904–1920, 3*(1–2), 94–177.

Burt, C. (1911). Experimental tests of higher mental processes and their relation to general intelligence. *Journal of Experimental Pedagogy, 1*, 93–112.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming.* Thousand Oaks, CA: Sage.

Canivez, G. L., & Kush, J. C. (2013). WAIS-IV and WISC-IV structural validity: Alternate methods, alternate results. Commentary on Weiss et al. (2013a) and Weiss et al. (2013b). *Journal of Psychoeducational Assessment, 31*(2), 157–169.

Carroll, J. B. (2003). The higher stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.). *The scientific study of general intelligence: Tribute to Arthur R. Jensen.* New York: Pergamon.

Clauset, A., Moore, C., & Newman, M. E. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature, 453*(7191), 98.

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality, 54*, 13–29.

Crowley, O. V., Kimhy, D., McKinley, P. S., Burg, M. M., Schwartz, J. E., Lachman, M. E., ... Sloan, R. (2016). Vagal Recovery From Cognitive Challenge Moderates Age-Related Deficits in Executive Functioning. *Res Aging, 38*(4), 504–525.

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2017). *A network perspective on political attitudes: Testing the connectivity hypothesis*.

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113*(4), 842.

van der Maas, H. L. J., Kan, K. J., Marsman, M., & Stevenson, C. E. (2017). Network models for cognitive development and intelligence. *Journal of Intelligence, 5*(2), 16.

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review, 108*(2), 346.

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*(1), 195–212.

Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48*(4), 1–18.

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network pschometrics: Combining network and latent variable models. *Psychometrika, 82*(4), 904–927.

Franić, S., Dolan, C. V., Borsboom, D., Hudziak, J. J., van Beijsterveldt, C. E., & Boomsma, D. I. (2013). Can genetics help psychometrics? Improving dimensionality assessment through genetic factor modeling. *Psychological methods, 18*(3), 406.

Fried, E. I., Bockting, C., Arjadi, R., Borsboom, D., Amshoff, M., Cramer, A. O., & Stroebe, M. (2015). From loss to loneliness: The relationship between bereavement and depressive symptoms. *Journal of Abnormal Psychology, 124*(2), 256.

Fried, E. I., van Borkulo, C. D., Cramer, A. O., Boschloo, L., Schoevers, R. A., & Borsboom, D. (2017). Mental disorders as networks of problems: A review of recent insights. *Social Psychiatry and Psychiatric Epidemiology, 52*(1), 1–10.

Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review, 109*(1), 116.

Geary, D. C. (2018). Efficiency of mitochondrial functioning as the fundamental biological mechanism of general intelligence (g). *Psychological Review, 125*(6), 1028–1050.

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research, 48*(5), 639–662.

Golino, H. F., & Demetriou, A. (2017). Estimating the dimensionality of intelligence like data using exploratory graph analysis. *Intelligence, 62*, 54–70.

Gottfredson, L. S. (2017). A *g* theorist on Why Kovacs and Conway's process overlap theory amplifies, not opposes, *g* theory. *Psychological Inquiry, 27*(3), 210–217.

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., ... Koellinger, P. D. (2018). *Genomic SEM provides insights into the multivariate genetic architecture of complex traits.* bioRxiv305029.

Gustafsson, J. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*(3), 179–203.

Hill, W. D., Marioni, R. E., Maghzian, O., Ritchie, S. J., Hagenaars, S. P., McIntosh, A. M., & Deary, I. J. (2018). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular Psychiatry*. https://doi.org/10.1038/s41380-017-0001-5 Epub ahead of print.

Hofman, A., Kievit, R. A., Stevenson, C., Molenaar, D., Visser, I., & van der Maas, H. L. J. (2018). *The dynamics of the development of mathematics skills: A comparison of theories of developing intelligence.* https://doi.org/10.31219/osf.io/xa2ft.

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2*(1), 41–54.

Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution.* Chicago: University of Chicago Press.

Hood, S. B. (2008). *Latent Variable Realism in Psychometrics.* Doctoral Dissertation*Department of History and Philosophy of Science*. Bloomington, IN: Indiana University.

Horn, J. L. (1998). A basis for research on age differences in cognitive abilities. In J. J. McArdle, & R. W. Woodcock (Eds.). *Human cognitive abilities in theory and practice* (pp. 57–92). Mahwah, NJ: Erlbaum.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Johnson, W., Bouchard, T. J., Jr., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence, 32*(1), 95–107.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32*(4), 443–482.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications.* SPSS.

Kievit, R. A., Lindenberger, U., Goodyer, I. M., Jones, P. B., Fonagy, P., Bullmore, E. T., ... Dolan, R. J. (2017). Mutualistic coupling between vocabulary and reasoning supports cognitive development during late adolescence and early adulthood. *Psychological Science, 28*(10), 1419–1431.

Kievit, R. A., Van Rooijen, H., Wicherts, J. M., Waldorp, L. J., Kan, K., Scholte, H. S., & Borsboom, D. (2012). Intelligence and the brain: A model-based approach. *Cognitive Neuroscience, 3*(2), 89–97.

Kline, R. B. (2015). *Principles and practice of structural equation modeling.* New York: Guilford publications.

Kovacs, K., & Conway, A. R. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry, 27*(3), 151–177.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence, 14*(4), 389–433.

Lachman, M. E., Agrigoroaei, S., Tun, P. A., & Weaver, S. L. (2014). Monitoring cognitive functioning: Psychometric properties of the brief test of adult cognition by telephone. *Assessment, 21*(4), 404–417.

Lachman, M. E., & Tun, P. A. (2008). Cognitive testing in large-scale surveys: Assessment by telephone. In S. M. Hofer, & D. F. Alwin (Eds.). *Handbook of cognitive aging: Interdisciplinary perspectives* (pp. 506–522). Thousand Oaks, CA: Sage.

Levine, S. Z., & Leucht, S. (2016). Identifying a system of predominant negative symptoms: Network analysis of three randomized clinical trials. *Schizophrenia Research, 178*, 17–22.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149.

Mackintosh, N. J. (2011). *IQ and human intelligence.* Oxford University Press.

Madhoo, M., & Levine, S. Z. (2016). Network analysis of the quick inventory of depressive symptomatology: Reanalysis of the STAR* D clinical trial. *European Neuropsychopharmacology, 26*(11), 1768–1774.

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. M. Collins, & A. G. Sayer (Eds.). *Decade of behavior. New methods for the analysis of change* (pp. 139–175). Washington, DC, US: American Psychological Association.

Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence, 3*(1), 2–20.

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41*(5), 407–422.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus statistical modeling software: Release 7.0.* Los Angeles, CA: Muthén & Muthén.

Neale, M. C., & Cardon, L. R. (1994). Methodology for genetic studies of twins and families. *Statistics in Medicine, 13*, 199.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., ... Sternberg, R. J. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*(2), 77.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer,

E. (2012). Intelligence: New findings and theoretical developments. *American psychologist, 67*(2), 130.

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics, 19*(3), 148.

Popper, K. (2005). *The Logic of Scientific Discovery*. London: Routledge.

R Core Team (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Raykov, T., & Marcoulides, G. A. (2012). *A first course in structural equation modeling.* New York: Lawrence Erlbaum.

Ripley, B. D. (2002). *Modern applied statistics with S. Statistics and Computing* (Fourth Ed). New York: Springer.

van Rooijen, G., Isvoranu, A., Kruijt, O. H., van Borkulo, C. D., Meijer, C. J., Wigman, J. T., ... Cahn, W. (2018). A state-independent network of depressive, negative and positive symptoms in male patients with schizophrenia spectrum disorders. *Schizophrenia Research, 193*, 232–239.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software, 48*(2), 1–36.

Ruzzano, L., Borsboom, D., & Geurts, H. M. (2015). Repetitive behaviors in autism and obsessive–compulsive disorder: New perspectives from a network analysis. *Journal of Autism and Developmental Disorders, 45*(1), 192–202.

Sachisthal, M. S., Jansen, B. R., Peetsma, T. T., Dalege, J., van der Maas, H. L. J., & Raijmakers, M. E. (2018). Introducing a science interest network model to reveal country differences. *Journal of Educational Psychology.*. https://doi.org/10.1037/edu0000327.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*(1), 53–61.

Schneider, W. J., & McGrew, K. S. (2018). The cattell-horn-Carroll theory of cognitive abilities. *Contemporary Intellectual Assessment: Theories, Tests and Issues,* 73–130.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*(3), 333–343.

Spearman, C. (1904). " General intelligence," objectively determined and measured. *The American Journal of Psychology, 15*(2), 201–292.

Spearman, C. (1927). *The abilities of man.* Oxford, England: Macmillan.

Taylor, P., Hobbs, J. N., Burroni, J., & Siegelmann, H. T. (2015). The global landscape of cognition: Hierarchical aggregation as an organizational principle of human cortical networks and functions. *Scientific Reports, 5*, 18112.

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review, 38*(5), 406.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Tobin, R. M. (2013). An introduction to the wechsler intelligence tests: Revisiting theory and practice. *Journal of Psychoeducational Assessment, 31*(2), 91–93.

Tun, P. A., & Lachman, M. E. (2006). Telephone assessment of cognitive function in adulthood: The brief test of adult cognition by telephone. *Age and Ageing, 35*(6), 629–632.

Tun, P. A., & Lachman, M. E. (2008). Age differences in reaction time and attention in a national telephone sample of adults: Education, sex, and task complexity matter. *Developmental Psychology, 44*(5), 1421.

Van der Sluis, S., Posthuma, D., & Dolan, C. V. (2013). TATES: Efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genetics, 9*(1), e1003235.

Ward, L. C., Bergman, M. A., & Hebert, K. R. (2012). WAIS-IV subtest covariance structure: Conceptual and statistical considerations. *Psychological Assessment, 24*(2), 328.

Wechsler, D. (2008). *Wechsler adult intelligence Scale–Fourth edition (WAIS–IV).* San Antonio, TX: The Psychological Corporation.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WAIS-IV and clinical validation of the four-and five-factor interpretative approaches. *Journal of Psychoeducational Assessment, 31*(2), 94–113.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics, 9*(1), 60–62.