

Simulating the Process Overlap Theory of Intelligence:  
A Unified Framework Bridging Psychometric and Cognitive Perspectives

Han Hao

Andrew R. A. Conway

New Mexico State University

Kristof Kovacs

ELTE Eotvos Lorand University

Jean-Paul Snijder

Heidelberg University

New Mexico State University  
2023

## Abstract

Traditional theories of intelligence either prioritize a psychometric or a cognitive perspective, but their limitations and incompatibilities hinder a comprehensive understanding. Contemporary theories, like the process overlap theory (POT; Kovacs & Conway, 2016; 2019), aim to bridge the gap between the two perspectives, by explaining inter-individual differences in intelligence through intra-individual psychological processes. The current study investigates POT as a unified framework for understanding human intelligence, incorporating psychometric and cognitive theories. POT proposes a novel psychometric structure and cognitive architecture that explains individual differences in cognitive abilities. We developed dynamics to simulate potential correlational/causal structures of cognitive processes involved in human cognitive activities based on POT, examining how these structures align with psychometric models. Test scores were generated from a sampling of simulated cognitive processes and fitted by typical latent factor models. Despite the absence of a general cognitive ability in generating the data, results showed that a standard higher-order "general intelligence" model fit the data well. As POT rejects the notion of a general factor of intelligence ( $g$ ), psychometric network models (Borsboom et al., 2021; Epskamp et al., 2018) were also implemented to simulated test scores, as they align better with the theory. Finally, we implemented a network structure at the latent factor level to retain latent variable models' benefits in accounting for measurement error. Estimated factor scores for simulated broad abilities from the three different models are compared and discussed. This study demonstrates POT's compatibility with standard psychometric models, including the general intelligence factor, without assuming a common cognitive cause. The results support POT and provide an alternative theoretical and statistical framework for

contemporary research on human cognition, combining psychometric and cognitive theories of intelligence.

Preprint not peer reviewed

## **Simulating the Process Overlap Theory of Intelligence:**

### **A Unified Framework Bridging Psychometric and Cognitive Perspectives**

Human intelligence has been a controversial topic in scientific research for over a century, with different fields proposing various operational definitions (Sternberg, 2013). Indeed, a widely accepted definition of intelligence remains elusive. Despite the lack of agreement on a definition, researchers studying intelligence have identified several robust findings in cognitive ability testing. The positive manifold is one such finding, where performance on subtests of different mental abilities is positively correlated (Carroll, 1993; Neisser et al., 1996; Spearman, 1927). This has led to the concept of a “general intelligence”, or *g*, which statistically, and sometimes also conceptually, represents this positive manifold and plays a central role in all intellectual activities. In some intelligence models, *g* is regarded as a higher-order factor of intelligence that accounts for the observed covariance among broad ability factors (Cattell, 1963; Horn, 1968; McGrew, 2005). The *g* factor accounts for 40-60% of the variance in intelligence tests (Deary et al., 2010; Jensen, 1998) and is predictive of a wide range of life outcomes, including academic achievement, job performance, and income (Gottfredson, 1997).

Traditional theories of intelligence have provided important insights to measure, explain, and predict individual differences in human intellectual behaviors, especially on the positive manifold and general intelligence. Psychometric theories of intelligence investigate the correlational relationship among behavioral performance in different cognitive activities (McGrew, 2009; Neisser et al., 1996). In psychometric theories, *g* is generally regarded as a general mental ability that causes performance on different ability tests to correlate. These theories originated in Spearman’s (1904) work on the factor analysis of school test scores and imply that *g* is a measure of a psychological attribute that is indirectly measured by the correlated

test scores. One of the most representative psychometric models of intelligence is the hierarchical CHC model, in which the structure of intelligence is described by a hierarchical multi-factor model. The CHC model has been influential in the construction of many current IQ tests such as the Stanford-Binet Intelligence Scale (5<sup>th</sup> Edition) and Woodcock-Johnson III and IV tests (Roid & Pomplun, 2012; Schrank, 2011; Schrank et al., 2016). It has been recognized as a comprehensive account of the correlational structure of diverse intelligence tests across test batteries (Conway & Kovacs, 2015; Flanagan et al., 2013).

Most traditional psychometric theories are based on a broad premise: common variance among different measures reflects a general source that causes the communality of the individual differences, with unique variances reflecting task/measure specific resources. This premise has led to a potential ambiguity between “psychometric *g*” (a between-subject construct that statistically accounts for the positive correlations between psychometric tests) and “psychological *g*” (a within-subject construct; a general mental ability that is involved in every kind of cognitive activity). Psychometric theories typically explain interpersonal individual differences well but are not as good at explaining intrapersonal individual differences, because most psychometric theories are based on task-level correlations, which are usually estimated from composite scores of items with invariant formats and properties within each task. These items are regarded as equivalent in a task and the variation among items in the same task is usually perceived as measurement error, as is the variation among tasks of a similar type.

Cognitive theories, on the other hand, focus on identifying specific cognitive processes associated with general intelligence or specific cognitive abilities (Sternberg, 1985). Contrary to most psychometric theories that attempt to explain complex intellectual activities with one or a few latent factors as the common cause, cognitive theories narrow down to simpler forms of

intellectual activities, in which specific cognitive processes/mechanisms are engaged as basic information-processing components (Hunt et al., 1975; Sternberg, 1985). Most cognitive theories have attempted to understand  $g$  by deconstructing intellectual behaviors into their basic information-processing components and finding the “most influential” cognitive processes/mechanisms of intelligence. For example,  $g$  has been equated with various constructs, such as information processing speed (Jensen, 1998), working memory capacity (Kyllonen & Christal, 1990), and attention control (Engle, 2002).

One of the fundamental challenges for the cognitive approach to intelligence is the measurement problem (Frischkorn & Schubert, 2018). Cognitive models are derived from empirical research using elementary cognitive tasks (ECTs) designed to reflect individual differences in specific cognitive processes. However, ECTs in the real world can hardly be “process pure” and may involve interactions with other confounding processes such as top-down strategies (Deary, 2003). Some analytic approaches have been used in cognitive research to address this problem. For instance, the “cognitive-correlates” approach estimates latent factor models from multiple cognitive measures on the same process (e.g., Hunt et al., 1975); and the “cognitive-components” approach compares subjects’ cognitive performances in different experimental conditions of a task, in which these conditions are designed to require different combinations of processes (Sternberg, 1985). However, the “cognitive-correlates” approach usually requires large batteries of tasks and relies on the same premises as the psychometric models, while the “cognitive-components” approach often requires many trials and multiple experimental conditions to estimate reliable intrapersonal individual differences.

Furthermore, despite the valuable insights provided by traditional psychometric and cognitive theories, these two approaches are hardly compatible with each other. Psychometric

theories offer comprehensive descriptions of the structure of intelligence, but they cannot specify the exact cognitive processes and operational mechanisms that prompt the correlational structures of intelligence. On the other hand, cognitive theories deconstruct cognitive behavior by isolating essential cognitive processes but are not good at explaining how these "isolated" cognitive processes cooperate with each other and contribute to intelligence as a system. This gap between the two approaches highlights the need for new intelligence research perspectives that can bridge the gap between psychometric and cognitive theories, with updated theoretical inference and advanced computational realization.

### **Process Overlap Theory**

Contemporary intelligence theories strive to provide a more unified perspective by combining psychometric and cognitive approaches (Conway & Kovacs, 2015; Kaufman et al., 2013; Savi et al., 2019). One recent framework, the process overlap theory (POT; Kovacs & Conway, 2016, 2019), offers an alternative explanation to a range of primary findings in psychometric and cognitive intelligence research. Drawing from the sampling explanation of general intelligence (Thomson, 1916, 1951; Thorndike, 1925; see also Bartholomew et al., 2009), POT attempts to quantitatively define the theoretical framework of intelligence by offering a new perspective on the cognitive foundation of the positive manifold of intelligence and Spearman's  $g$ .

The sampling model of  $g$ , originally called the bonds model (Thomson, 1916), suggests that multiple mental "bonds" (which we now interpret as cognitive processes) are involved in the performance of intelligence tests and are sampled across a battery of tests. By this account,  $g$  is not the cause of the all-positive correlations between psychometric tests. Rather, it is the consequence of multiple bonds being sampled in an overlapping manner across a battery of

tasks. While the bonds model offered an alternative explanation of  $g$ , it also had a number of shortcomings that contributed to its diminished popularity (Eysenck, 1987; van der Maas et al., 2006).

Similar to the initial sampling models of  $g$ , POT also regards the positive manifold as the consequence of overlapping psychological resources and exhibits a computational framework of intelligence that does not rely on the latent common cause assumption of factor models. Furthermore, POT employs a sampling mechanism for different types of cognitive processes, where domain-specific and domain-general processes are sampled in an overlapping manner across a battery of tests (Kovacs & Conway, 2016). It is argued that the POT sampling mechanism, along with the specification of different types of cognitive processes, can capture the hierarchical structure of intelligence observed in traditional psychometric models of intelligence. This contemporary framework, motivated by both psychometric and cognitive theories, serves as a bridge between these traditionally separated approaches. It rejects the notion of a psychological  $g$  in intelligence and can better harmonize intelligence research evidence found in psychometrics, cognitive psychology, and neuroscience.

According to Kovacs & Conway (2019), process overlap theory is portrayed at four different levels of representation: (a) the conceptual model that is verbally described (POT-V), (b) the multi-dimensional item response model (POT-I) that provides a computational framework based on item response theory, in which lower-level cognitive processes are sampled and reflected in higher-level behavior performance, (c) the structural model (POT-S) that represents the intercorrelations of higher-level behavior performance based on the sampling algorithm in a traditional latent factor structure, and (d) the psychometric network model (POT-N) that depicts



the same intercorrelations, but in a psychometric network that does not rely on the assumption of latent common causes of task performance.

### ***The Conceptual Model (POT-V)***

Process overlap theory extends the initial sampling theories by proposing a nonadditive sampling mechanism of cognitive processes. Specifically, in POT, intelligence tests in any battery tap into a number of domain-general processes and domain-specific processes. While the domain-general processes, such as the cognitive processes involved in working memory, are sampled across all tests, the domain-specific ones are sampled only in some of them. In this POT sampling mechanism, there is no single process that is required by all items and tests. Instead, some processes, the domain-general ones, are sampled more frequently than the domain-specific processes. Consequently, the positive correlations observed on various intelligence tests stem from overlapping cognitive processes tapped by different tests. Furthermore, the different domains tapped by the given task are modeled as ‘sub-tasks’ that have to be solved independently, hence while within-domain processes compensate for each other, cross-domain processes do not. Therefore, the limits posed by individual differences in domain-general processes will function as a bottleneck to overall performance (Kovacs & Conway, 2019).

### ***The Multidimensional Item Response Model (POT-I)***

This conceptual model of POT is represented in a multi-dimensional item response (mIRT) model (POT-I) described by Equation 1:

$$P(U_{pi} = 1 | \theta_{plm}, a_{il}, b_{il}) = \prod_{l=1}^D \frac{e^{\sum_{m=1}^C a_{il}(\theta_{plm} - b_{il})}}{1 + e^{\sum_{m=1}^C a_{il}(\theta_{plm} - b_{il})}} \quad \#(1)$$

In Equation 1, the probability ( $P$ ) of an individual ( $p$ ) correctly answering a test item ( $i$ ) depends on their ability level ( $\theta$ ) on the processes required by that item, along with the item

characteristics, represented by the item discrimination and difficulty parameters.  $\theta_{plm}$  represents the level of ability for the  $p$ -th individual on the  $m$ -th process in the  $l$ -th domain;  $a_{il}$  is the discrimination parameter for the  $i$ -th item in the  $l$ -th domain;  $b_{il}$  is the difficulty parameter for the  $i$ -th item in the  $l$ -th domain.  $D$  is the number of domains sampled by an item and  $C$  is the number of processes in a given domain sampled by the item.

In the POT-I algorithm, the cognitive processes are compensatory within domains but non-compensatory across domains. These cognitive processes constitute the assumed latent traits, or the ability levels, required for specific items. These cognitive processes are sampled and composed in two compensatory manners: additive or non-additive. For a specific item response, the ability levels ( $\theta$ ) on the processes within the same domain are additive, but each domain of a test item functions as a separate dimension, and each of the dimensions has to be passed independently for the item to be solved correctly (reflected by the multiplication function in Equation 1). For example, if a visuospatial test requires processes in two domains, executive and spatial, then a weakness in a spatial process can be compensated by another spatial process, but such compensation cannot take place across domains. Thus, according to POT-I, for a specific item, the observed performance reflects a function of multiple domain-general abilities and multiple domain-specific abilities and these ability levels are expressed in Equation 1 as  $\theta_{plm}$ .

### ***The Structural Model (POT-S)***

According to POT, the multidimensional item response model described in POT-I can result in positive task-level intercorrelations among different cognitive tasks without any single cognitive process being sampled across all tasks. The patterns of these intercorrelations are expected to exhibit a positive manifold similar to the positive manifold observed in conventional psychometric theories. Traditional psychometric theories apply exploratory or confirmatory

factor analyses to these correlations to find the common factors. These common factors are reflective and are interpreted as broad cognitive abilities that function as the common causes of task correlations (Borsboom et al., 2003). Similar to traditional psychometric theories, POT predicts a higher-order factor structure of psychometric intelligence. However, POT interprets the commonly observed higher-order  $g$  factor in traditional psychometric models of intelligence as a formative factor that represents the common consequence of the positive manifold, instead of a reflective factor that represents the common cause (Kovacs & Conway, 2016, 2019). Thus, in POT-S the direction of causality between broad ability factors and  $g$  is opposite to that in traditional models like the CHC theory. The opposite direction of causality indicates that psychometric  $g$  is represented as a weighted combination of lower-level broad abilities as indicators. The psychological interpretation of this formative higher-order factor is dependent on the selection of indicators (intelligence tests). In other words, individual differences in  $g$  are defined and determined by the cognitive measures included in a particular psychometric model of intelligence.

This formative model adopts a more conservative perspective on the causal inference of  $g$ , such that the formative  $g$  does not rely on the common cause assumption. Reminiscent of Boring (1923), the formative model defines intelligence as what the intelligence tests test. Thus, POT-S illustrates a factor structure of psychometric intelligence in which the subfactors, representing broad abilities measured by different tests, are reflective latent factors, while the higher-order  $g$  is a formative factor emerged from the subfactor correlations, which are caused by the sampling mechanisms proposed in POT-I.

### ***The Network Model (POT-N)***

Due to the common cause assumption, the latent factor approach is not compatible with the theoretical perspective of POT. As previously mentioned, POT interprets  $g$  as a result of the sampling mechanism on cognitive processes. Importantly, none of the cognitive processes is exclusively involved across an entire battery of tests. Therefore, from a theoretical perspective, a network model (Kan et al., 2019) is more compatible with POT-V at the task level than a latent variable model, such that the  $g$  factor is not assumed to be the cause of the covariance among intelligence tests but is caused by an overlap of processes tapped by tests. Recent developments in the psychometric network approach (Borsboom et al., 2021; Epskamp et al., 2018; van der Maas et al., 2017) have made it possible to estimate conditional dependency among the tests as well as assess the stability and accuracy of the network structures, without assuming latent common causes.

Network models, as they are titled, represent the covariance structure among observed variables (in the form of nodes) by a partial correlation network (in the form of undirected weighted edges). Partial correlations reflect the remaining association between two nodes after controlling for all other information available (Epskamp & Fried, 2018). Thus, the covariance structure among tests is represented in a network structure where the observed tests are the nodes and their pairwise partial correlations are the edges. POT-N proposes that the broad abilities are reflected in small “clusters” of nodes, connected by strong edges among the test scores that are the same test type. Therefore, POT-N does not assume any latent factors to represent the covariance structure.

However, it is worth mentioning that the psychometric network model presented here in POT-N does not indicate that a network model can validate POT by itself (Conway et al., 2020).

For a set of behavioral observations of intelligence (test scores), a network model, with nodes as subjects' performance on specific tests and edges as their relationships, is compatible with POT-V, whereas a reflective latent factor model, with manifest variables representing test performance and factor loadings representing their reliabilities of reflecting the latent construct (intelligence), is incompatible with POT-V. It is important to note here that the current POT-N is not fully representing POT-V because it does not represent the underlying cognitive mechanisms that POT-V describes. This is because the behavioral observations (test scores) are not direct representations of the sampled cognitive processes, and the between-test conditional relationships revealed by the network model do not reveal the between-process conditional relationships. The current study is designed to compare latent variable models and network models, so for simplicity, we specified test scores as the observed variables. A goal for future research is to obtain more direct measures of cognitive processes based on computational cognitive models of task performance.

### **The Current Study**

Process overlap theory attempts to explain intelligence and corresponding major phenomena from a new perspective and with a computational framework that could, to some extent, unify the traditional psychometric and cognitive models (Kovacs & Conway, 2019). Specifically, POT proposes information-processing mechanisms from a cognitive perspective and maps the cognitive mechanisms into correlational structures of latent ability factors proposed by psychometric theories.

The current study directly applies the Process Overlap Theory (POT) to investigate the cognitive and psychometric structure of intelligence with two primary aims. First, we want to demonstrate that, according to POT-I, the positive manifold, and therefore  $g$ , can emerge in the

absence of a psychological common cause (Conway et al., 2021). Furthermore, with the distinction between domain-general and domain-specific processes, we aim to demonstrate that a hierarchical structure of intelligence can emerge (Conway et al., 2021). Therefore, in Study 1 we conducted a simulation that applied the POT-I algorithm to generate test scores from a sampling of simulated cognitive processes and then fitted the simulated test scores with typical latent factor models. This simulation provides a “proof of principle”, to conceptually illustrate the sampling mechanism of POT at both the cognitive level and the psychometric level.

The second aim of the current study is to extend the current factor model of intelligence (POT-S) to an alternative network structure (POT-N) that is more compatible with POT-V. Thus, in Study 2 we applied undirected network models to the simulated data from Study 1, which provides an alternative representation of the positive manifold among cognitive test scores. The network structure emphasizes direct mutual associations among cognitive measures, instead of an unobservable common cause, and is, therefore, more compatible with POT-V. Furthermore, a latent network approach (Epskamp et al., 2017) was applied to the simulated data, in which the benefits of latent factor models and psychometric network models were combined. The latent network model accounts for the concern of measurement error and retains the properties of latent variable models at the broad ability level, and represents the conditional dependence of different broad abilities in the form of an undirected network. Finally, factor scores (and their equivalent statistics in network models) were estimated and compared for the simulated broad abilities from the latent factor model, the network model, and the latent network model.

### **Study 1: A Simulation Framework Based on POT**

The simulation was conducted to translate the conceptual and IRT model of POT (POT-V and POT-I, respectively) to sampling algorithms on randomly generated matrices, which

represent subjects' individual differences in all the cognitive processes that are relevant to designated cognitive tests. In this simulation, subject scores on a battery of intelligence tests were calculated as a function of domain-general and domain-specific ability parameters. Test scores were simulated according to POT-V and POT-I and were compared to a generic sampling model (GSM) in which all the cognitive processes are assumed to be equally weighted across the battery of tests. The GSM approach contrasts with POT such that GSM represents the lack of a distinction between general and specific cognitive processes. R Scripts for the simulation are available at <https://osf.io/jhzyyp/>.

For simplicity, the simulation specified three broad cognitive abilities: fluid, verbal, and spatial. Each of the broad abilities was assumed to be measured by 3 tests. Each test was assumed to consist of 100 items. To obtain a psychometric model of intelligence with a higher-order  $g$  factor, a sample of 1000 subjects was simulated. Individual subject scores for the 9 tests were generated based on the POT-I algorithm and a GSM algorithm. Item scores (correct or incorrect; represented by 1 or 0, respectively) were determined by both domain-specific ability and domain-general ability in the two algorithms, which were represented by functions of simulated processes sampled by the item.

This simulation consisted of four steps: (a) specification of the simulated cognitive processes and tests; (b) calculation of ability parameters ( $\theta_s$ ) based on sampling algorithms; (c) calculation of item scores; and (d) test scoring and psychometric model fitting.

### **Step 1: Specification of Cognitive Processes and Tests**

For each of the 1000 subjects, 60 cognitive processes were specified. The 60 cognitive processes were classified into four domains (15 per domain): fluid reasoning processes, verbal processes, spatial processes, and executive function processes. These are the four domains that

are assumed to be involved in the 9 tests measuring 3 broad abilities (fluid, verbal, and spatial). Each subject was thought to have an “ability level” on each of the 60 cognitive processes. These cognitive process abilities were represented by numerical values from a standardized multivariate normal distribution,  $X_p \sim \mathcal{N}_{200} (\mu = 0, \Sigma = I)$ . This simulation step resulted in a  $1000 \times 60$  matrix (number of subjects  $\times$  number of processes) of orthogonal process ability values.

## Step 2: Calculation of Ability Parameters ( $\Theta$ s) Based on Two Algorithms

As aforementioned, the POT model was inspired by the initial sampling theory by Thomson (1916). A simple elaboration of sampling theory is that the ability,  $y$  (in this simulation,  $\Theta$ ), required for an individual on a specific item of a test can be expressed as

$$y = \mathbf{a}^T \mathbf{x} = \sum_{k=1}^n a_k x_k \quad \#(2)$$

In Equation 2,  $n$  is the number of simulated cognitive processes in the corresponding domain,  $\mathbf{a}^T$  is an  $n$ -element row vector of random binominal values (0s and 1s) for process sampling, and  $\mathbf{x}$  is an  $n$ -element column vector that includes the  $n$  simulated cognitive processes in this domain for the individual. Therefore, the vector  $\mathbf{a}$  expresses how potential processes of the domain are involved for the individual to respond to the item, with  $a_k = 0$ , meaning process  $x_i$  is not sampled, and  $a_k = 1$ , meaning process  $x_k$  is sampled in achieving the corresponding level of ability for the item. As follows, the ratio of 0s and 1s in  $\mathbf{a}$  vector as the “sampling vector” determines the number of processes being sampled for a  $\Theta$  value. Both the POT algorithm and the GSM algorithm follow this basic sampling dynamic in this simulation. Details for both algorithms are described as follows.

### **POT Algorithm**

In the POT algorithm, each of the nine broad ability tests sampled processes from two of the four different domains: one of the three types of domain-specific processes (fluid, verbal, and



spatial processes for fluid, verbal, and spatial tests, respectively) plus the domain-general processes (executive function processes). While EF processes were allowed to be sampled in all three types of tests (fluid, verbal, and spatial), the other three types of processes could only be involved in their corresponding test. Hence, for each of the 100 item-level responses of an individual in a test, two aggregate process ability parameters (two  $\Theta$ s) were calculated: one domain-general and one domain-specific. The mathematical representation of the POT algorithm in Step 2 is described by the following equations (Equations 3 and 4):

$$\theta_{g_{pti}} = \sum_{n=1}^{15} b_{tin} G_{pn} \#(3)$$

$$\theta_{s_{pti}} = \sum_{m=1}^{15} c_{tim} S_{pm} \#(4)$$

In these two equations,  $\theta_{g_{pti}}$  and  $\theta_{s_{pti}}$  are the aggregate domain-general and aggregate domain-specific ability parameters for the  $p$ -th individual on the  $i$ -th item of the  $t$ -th test. The aggregate ability parameters are the sums of sampled processes from their corresponding domains (G as the domain-general EF processes and S as the domain-specific processes) based on the sampling vectors  $b_{ti}$  and  $c_{ti}$ .

An assumption of POT is that all cognitive tests sample domain-general processes and the probability of a domain-general process being sampled is higher in fluid reasoning tests than in other tests (Kovacs & Conway, 2016). In this simulation, for an item from a fluid reasoning test, the probability of an EF process being sampled was set to 0.28, while the probability of a fluid-specific process being sampled was 0.12, which on average sample about four EF processes and two fluid processes for an item on a fluid reasoning test. However, for an item from a domain-specific test (verbal or spatial), the probability of an EF process being sampled was set to 0.12, while the probability of a domain-specific process (verbal or spatial, respectively) being

sampled was 0.28, which on average sample about two EF processes and four verbal/spatial processes for a domain-specific item. These values of sampling probability were arbitrarily selected, such that in either type of tests, there were on average a total of 6 processes sampled when generating a simulated response to an item.

Finally, two  $1000 \times 9 \times 100$  (number of subjects  $\times$  number of tests  $\times$  number of items per test) three-dimensional arrays of aggregate ability parameters ( $\theta_g$  and  $\theta_s$ ) were created as the result of this step for the 9 tests, representing the domain-general and domain-specific dimensions of traits required to respond correctly on the test items.

### ***GSM Algorithm***

For the generic sampling algorithm, no domains were specified when sampling the cognitive processes across the 9 tests, and thus only one aggregate ability parameter ( $\theta$ ) was calculated for each of the 100 item-level responses for an individual on a test. Therefore, when generating simulated item-level responses for each test, the probabilities of the 60 cognitive processes being sampled for a corresponding process parameter ( $\theta$ ) were all equal, as reflected in Equation 5:

$$\theta_{pti} = \sum_{h=1}^{60} d_{tih} E_{ph} \quad \#(5)$$

In this equation,  $\theta_{pti}$  is the aggregate ability parameter for the  $p$ -th individual on the  $i$ -th item of the  $t$ -th test. Each aggregate ability parameter is the sum of sampled processes from the 60 processes of an individual based on the sampling vector  $d_{ti}$ .

In this simulation, the probability of a cognitive process being sampled was set to 0.10. Thus, in concordance with the POT algorithm, an average of 6 processes were sampled when generating a simulated response to an item. One  $1000 \times 9 \times 100$  (number of subjects  $\times$  number of

tests  $\times$  number of items per test) 3-dimensional array of aggregate ability parameters ( $\Theta$ ) was created as the result of this step for the GSM algorithm.

### Step 3: Calculation of Item Scores

To generate item-level scores based on the POT algorithm and the GSM algorithm, different IRT functions were applied to the ability parameters generated from the two algorithms.

#### *POT Algorithm*

For the POT algorithm, a multidimensional IRT function was applied to the domain-general and domain-specific ability parameters ( $\Theta_g$  and  $\Theta_s$ ). This function is a practical version of the conceptual POT-I model described in Equation (1). The mathematical representation is described by Equation (6).

$$P_{pti} = \frac{1}{1 + e^{-(z(\Theta_{gpti}) - b_{gti})}} \cdot \frac{1}{1 + e^{-(z(\Theta_{spti}) - b_{sti})}} \quad \#(6)$$

where  $P_{pti}$  is the probability of the  $p$ -th subject getting the  $i$ -th item in the  $t$ -th test correct;  $z(\Theta_{gpti})$  and  $z(\Theta_{spti})$  are the standardized scores of  $\Theta_{gpti}$  and  $\Theta_{spti}$ , respectively. In Equation (6), although processes are compensatory within a domain,  $\Theta_g$  and  $\Theta_s$  are not compensatory with each other.

In this simulation, the discrimination parameters ( $a$  parameters) were set to 1 for parsimony across the 100 items and 9 tests. The difficulty parameters of the domain-general and the domain-specific processes ( $b$  parameters,  $b_g$  and  $b_s$ ) for the 900 test items were randomly drawn from a standardized multivariate normal distribution, with each of the nine tests having average  $b_g$  and  $b_s$  of 0 and standard deviation of 1 and being independent of each other. Under Equation 6, the two three-dimensional arrays of  $\Theta_g$  and  $\Theta_s$  were transferred to one  $1000 \times 9 \times 100$  (number of subjects  $\times$  number of tests  $\times$  number of items per test) three-dimensional array of probabilities. A simulated binary response (0 for incorrect response and 1 for correct response) on each item of each test for each individual was generated on the basis of the probabilities from

the array. Thus, in general, a higher ability would generate a higher probability value to achieve the correct response to the item in a simulation.

### **GSM Algorithm**

For the GSM algorithm, a unidimensional IRT function was applied to the sampled process values ( $\theta e$ ) because no domains are specified, and hence, only one ability parameter for each simulated response was calculated. The mathematical representation is described by Equation 7:

$$P_{pti} = \frac{1}{1 + e^{-\left(z(\theta e_{pti}) - \frac{bg_{ti} + bs_{ti}}{2}\right)}} \#(7)$$

Similar to the POT algorithm,  $P_{pti}$  is the probability of the  $p$ -th subject getting the  $i$ -th item in the  $t$ -th test correct, and  $z(\theta e_{pti})$  is the standardized score of  $\theta e_{pti}$ .

All discrimination parameters ( $a$  parameters) were also set to 1 across the 100 items and 9 tests to maintain parsimony. The same sets of difficulty parameters of the domain-general and the domain-specific aspects ( $b$  parameters,  $bg$  and  $bs$ ) were applied to the GSM algorithm, but unlike the POT algorithm, for each simulated response, the average of the corresponding  $bg$  and  $bs$  was taken as the  $b$  parameter for that item. According to this equation, the 3-dimensional array of  $\theta e$  was transferred to a  $1000 \times 9 \times 100$  (number of subjects  $\times$  number of tests  $\times$  number of items per test) 3-dimensional array of probabilities. Similar to the POT algorithm, the 3-dimensional array of probabilities was applied to generate binary responses under the GSM algorithm.

### **Step 4: Test Scoring and Statistical Model Fitting**

For both algorithms, Step 3 generated a  $1000 \times 9 \times 100$  (number of subjects  $\times$  number of tests  $\times$  number of items per test) three-dimensional array of binary values (1 and 0), in which 1

represented a correct response and 0 represents an incorrect response. Thus, the simulated item-level responses were aggregated per individual, which results in a  $1000 \times 9$  (number of subjects  $\times$  number of tests) test-level dataset for each algorithm. In the two datasets, each subject had one score ranging from 0 to 100 for each of the nine tests, simulated under corresponding algorithms.

Two confirmatory factor models were applied to the two datasets to investigate theoretical latent factor structures that could best represent the data. First, a one-factor model was applied to both datasets, in which all 9 tests (as manifest variables) load onto the same  $g$  factor. Next, a three-factor model was applied, in which the 9 tests load onto three subfactors (fluid, verbal, and spatial) and the subfactors load onto a higher-order  $g$ . These two models were selected because the one-factor model was representative of Spearman's initial theory of general intelligence as well as Thomson's bonds model, whereas the higher-order three-factor model was representative of the current psychometric theories of intelligence, including the Cattell-Horn-Carroll model.

The fit of each model was evaluated with the following set of test statistics and fit indices:  $\chi^2$ , Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), and Akaike Information Criterion (AIC). The criteria for an "acceptable" model fit were based on the following cut-off values, recommended by Schreiber et al. (2006) and Kline (2015):  $p > .05$ ,  $\chi^2/df \leq 3$ ,  $CFI \geq .95$ ,  $RMSEA \leq .06$ ,  $SRMR \leq .08$ . For AIC there was no cut-off value; lower values indicate better fit.

### **The Simulation Procedure**

In the current study, the simulation process described above was conducted for 200 iterations. Factor loadings and fit indices from these 200 iterations were obtained and investigated. Parameters for these model parameters and indices were summarized and

compared. All steps of the simulation were conducted in R (version 4.1.1). The confirmatory factor analyses were conducted using the package “lavaan” (Rosseel, 2012).

The prediction of the results was based on the simulated cognitive algorithm of POT and GSM: for the POT algorithm, the one-factor model would convey poor model fit and the three-factor higher-order model would exhibit good fit. On the other hand, for the GSM algorithm, the one-factor model would convey good fit and the three-factor higher-order model would exhibit an overfit due to its complexity in factor structure. Furthermore, under the POT algorithm, the higher-order model would indicate factor patterns that are consistent with real-world observations, such that the higher-order factor loading for the fluid factor will be stronger than the verbal and spatial factors. Under the GSM algorithm, factor loadings would be all balanced across tests and broad ability factors, due to the lack of distinction between domain-general and domain-specific processes.

### Study 1 Results

The simulated data and all model fit statistics ( $\chi^2$  statistics, CFI, RMSEA, and SRMR) and factor loadings of the 200 iterations in the current study were available as data files at: <https://osf.io/jhzyp/>.

Table 1 presented the descriptive statistics for the fit indices from the four sets of models. The results suggest that, under the POT algorithm, the higher-order model exhibited a good fit for the simulated data whereas the one-factor model exhibited a poor fit in general. For the higher-order model,  $\chi^2$  statistics were all acceptably small with an average of 26.40, ranging from 11.36 to 47.35. Whereas for the one-factor model, the average  $\chi^2$  statistics was 1637.46, ranging from 1392.91 to 1858.29. For reference, the critical value for a significant  $\chi^2$  ( $df = 24$ , for the higher-order model,  $\alpha = .05$ ) is 36.42, and the critical value for a significant  $\chi^2$  ( $df = 27$ , for

the one-factor mode,  $\alpha = .05$ ) is 40.11. Results of the fit indices also support that the higher-order model fits the simulated datasets under the POT algorithm better than the one-factor model.

**Table 1.** Means (Standard Deviations) for the Fit Indices from the Four Sets of Models (by the Algorithm and the Model Structure)

		$\chi^2$	CFI	RMSEA	SRMR	AIC	BIC
POT	Higher-Order	26.21	1.00	.01	.02	58565.72	58668.78
		(6.57)	(< .01)	(.01)	(<.01)	(138.62)	(138.62)
	One-Factor	1637.46	.56	.24	.15	60170.97	60259.31
		(87.18)	(.02)	(.01)	(.01)	(169.36)	(169.36)
GSM	Higher-Order	26.77	.99	.01	.01	56411.03	56514.09
		(7.49)	(< .01)	(.01)	(<.01)	(127.53)	(127.53)
	One-Factor	27.69	1.00	.01	.01	56405.96	56494.30
		(7.65)	(< .01)	(.01)	(<.01)	(127.42)	(127.42)

*Note.*  $\chi^2$  = Chi-Squared Statistics. CFI = comparative fit index. RMSEA = root mean square error of approximation. SRMR = standardized root mean square residual. AIC = Akaike information criterion. The degrees of freedom for the higher-order models were 24. The degrees of freedom for the one-factor models were 27.

In contrast, under the GSM algorithm, both the higher-order model and the one-factor model provided an excellent fit for the simulated data. As Table 1 presents, model fit statistics for the one-factor model indicated excellent fit: the average  $\chi^2$  statistics was 27.69, ranging from 9.80 to 50.39; all indices indicated almost perfect fit. For the higher-order model, the average  $\chi^2$  statistics was 26.77, ranging from 9.39 to 48.78; and the indices also indicated almost perfect fit. However, the results of the higher-order model also indicated a problem of overfitting. In some of the fitted models, the higher-order loadings were approximately equal to 1 due to the high intercorrelations among the subfactors. In fact, for most (163 out of 200) simulated datasets

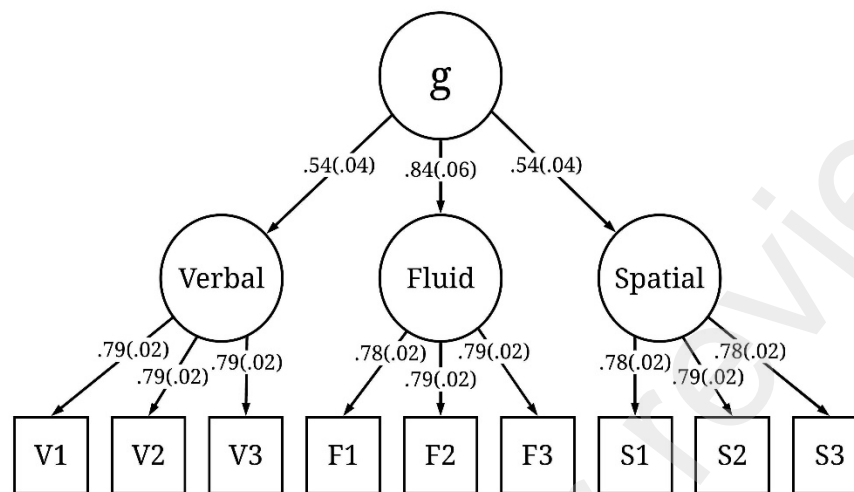
under the GSM algorithm, the model fitting results indicated that the higher-order model was an overfit to the simulated datasets. Given that the one-factor model has already fitted the data almost perfectly, it was clear that the higher-order structure was an over-specification of the GSM data. Thus, for the GSM algorithm, the one-factor model was the preferred model due to its more parsimonious nature. This was also supported by the comparison of the AIC and BIC indices from the two types of models, such that the one-factor models had slightly lower AICs and BICs.

Based on the results from these CFAs, Figure 1 presents means and standard deviations of the standardized factor loadings across the 200 iterations for the higher-order model of the POT algorithm. Figure 2 presents the same summary of the standardized factor loadings for the one-factor model of the GSM algorithm. The results were consistent with our prediction based on the two algorithms. For the POT algorithm, factor loadings for the manifest variables (representing the tasks) under their corresponding subfactors were balanced and stable across the 200 iterations. Furthermore, the higher-order factor loading for the fluid factor was stronger than the verbal and spatial factors. For the GSM algorithm, factor loadings for the 9 manifest variables were all balanced in the one-factor model and were stable across the 200 iterations.

### **Figure 1**



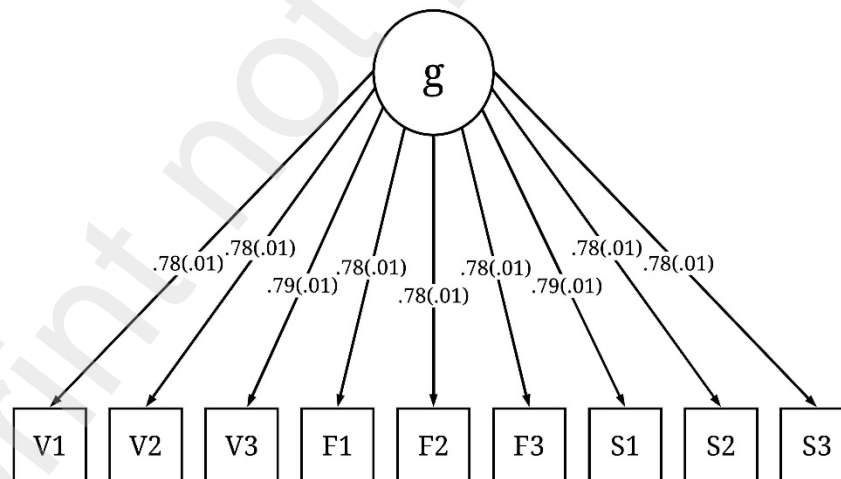
*The Higher-Order Model of Intelligence Based on 200 Iterations of Simulated Test Scores Under POT Algorithm*



*Note.* Values represent the means (and standard deviations) of standardized factor loadings. All parameters were estimated from simulated data.

**Figure 2**

*The One-Factor Model of Intelligence Based on 200 Iterations of Simulated Test Scores Under GSM Algorithm*



*Note.* Values represent the means (and standard deviations) of standardized factor loadings. All parameters were estimated from simulated data.

## Study 1 Discussion

The simulation in Study 1 is a conceptual demonstration that the positive manifold of intelligence (the statistical/psychometric  $g$ ) can emerge from the covariance of test scores in the absence of a general cognitive ability (a psychological  $g$ ). It also provides evidence in support of contemporary intelligence theories, including process overlap theory, offering a new alternative theoretical framework that unifies psychometric theories and cognitive theories of intelligence.

Two simulation algorithms (POT and GSM), both based on cognitive theories, were designed and described in Study 1. The GSM reflected a generic sampling method that was based on the traditional bonds model (Thomson, 1916). The POT algorithm was based on the process overlap theory, which was inspired by the traditional bonds model but proposed several extra central claims of the cognitive mechanisms underlying the sampling of cognitive processes. Both algorithms were specifically designed to avoid any simulated cognitive processes being sampled across all items or tasks. The POT algorithm contrasted with the GSM algorithm such that the POT algorithm specified a distinction between general and specific cognitive processes and how they were sampled in different cognitive tasks based on cognitive theories.

Simulated test scores were generated from these two algorithms and were fitted onto two CFA models. Across multiple iterations, the test scores generated from these two algorithms consistently reflected a positive manifold among 9 simulated tasks (3 fluid tasks, 3 verbal tasks, and 3 spatial tasks) without a “general” cognitive process as the common source of variance. The CFA results of these simulated test scores were consistent with corresponding psychometric theories. As was predicted by traditional bonds models, a one-factor model (conceptually similar to Spearman’s original model) fit the simulated data based on the GSM algorithm well, even in the absence of a common cognitive process across tests. The POT algorithm took a step further

by specifying different types of cognitive processes in the simulation algorithm. Based on the types of these hypothetical cognitive tests, the simulated cognitive processes were sampled in different manners per the corresponding cognitive properties of the tests. This specification as a simulated cognitive mechanism was reflected in the POT results. For the simulated data from the POT algorithm, a one-factor model was no longer appropriate and resulted in an unacceptable fit, whereas a higher-order three-factor model that was similar to the CHC model provided a great fit. Moreover, the POT algorithm also replicated a common psychometric finding that the fluid ability subfactor has a stronger loading on the higher-order  $g$  factor compared to other broad-ability subfactors in the higher-order model (Carroll, 1993).

It is worth noting that in the current study, all factor models were reflective. In the POT algorithm, a higher-order reflective latent factor model exhibited an excellent fit to the simulated data. However, we know that based on the simulation algorithm, not a single cognitive process was sampled across all tests as a “general mechanism”. In reality, psychometric  $g$  models are typically based on similar reflective factor models, such that the latent factor is regarded as a general causal mechanism across all cognitive tests. However, the current simulation demonstrates that a latent factor does not guarantee a common cognitive cause, even if the factor model fits the data well.

Overall, the simulation results support POT as a unified theory that attempts to bridge the gap between traditional cognitive theories and psychometric approaches. The POT algorithm proposed a conceptual illustration of the cognitive mechanisms in process overlap theory, which is an alternative interpretation of  $g$ , and the modeling results from this algorithm were consistent with findings in contemporary psychometric theories such as the CHC model. Although these results do not guarantee POT as the explanation of the inexistence of psychological  $g$ , they do

support the idea that a psychometrical/statistical  $g$  could be derived from a systematic sampling mechanism as was proposed by POT without a common source of variance to different types of cognitive tests.

### **Study 2: Psychometric Network Modeling of the POT Simulation**

In Study 1, the simulated test scores based on the POT algorithm and GSM algorithm were examined by latent factor models that are similar to conventional psychometric models. The latent factor models of the simulated data were applied to illustrate that a common cause was not a necessary condition for a common factor model to fit the data. However, according to POT-S, the  $g$  factor is a formative construct and does not assume a psychological general intelligence. Therefore, POT proposes a network structure of intelligence (POT-N) to represent the positive manifold among cognitive tests. For latent factor models, the covariance among observed variables is assumed to be caused by a common source that is represented by a reflective latent variable,  $g$ . However, for network models, complex systems of the observed variables are represented as interconnected networks. The psychometric network models regard the manifest variables as nodes in interconnected networks and their pairwise interactive associations as edges (Epskamp & Fried, 2018). Conceptually, these partial correlation edges among nodes are similar to undirected alternatives to the directed regression paths in regression models, except that there is no distinction between a “predictor” or an “outcome”. Thus, the partial correlations could be used to describe massive multicollinear relationships among cognitive testing data and transform the positive manifold into a graphic network. Hence, the psychometric network models are more compatible with process overlap theory than traditional latent factor models. Study 2 applied a psychometric network analysis on simulated data based

on the POT algorithm from Study 1 to provide an alternative representation (POT-N) of the intercorrelations among the simulated test scores.

Furthermore, another goal of Study 2 was to extend the standard psychometric network analysis approach by combining latent factor models and network models. Specifically, the simulated data were also investigated using a latent network model (Epskamp et al., 2017). POT rejects the notion of a psychological *g* but not the notion of broad abilities. However, POT-N in a standard psychometric network rejects not only the notion of general intelligence but also the notion of all broad abilities, by describing the common variance among all tests as an interconnected network. A better POT-N should exhibit psychometric networks at the broad ability level, but retain the latent factor structures at the lower level of specific task measures. In a latent network model, confirmatory hypotheses regarding different types of tasks are specified in a similar way to the hypotheses in a measurement model, and relationships among these latent factors are estimated in a similar way to those in a standard psychometric network model. Thus, the nodes in latent network analysis are latent factors based on prior theories and account for the measurement error in specific tasks, while the edges are estimated conditional dependence relationships among these latent factors. In Study 2, we referred to these nodes as “latent nodes” to reflect their conceptual similarity with the latent factors in factor models. A latent network model of POT, therefore, bridges POT-S and POT-N by accounting for measurement error in cognitive tasks while assuming no psychological *g*.

## **Study 2 Method**

The simulation process for the current study replicated the process in Study 1 but with a few updates. First, in the current study, only the POT algorithm was included in the simulation. Second, due to model convergence issues because of high correlations among simulated data, a

lower discrimination parameter of  $a = .70$  was introduced to this simulation. Third, an iteration of the simulation process was not conducted. The simulation in the current study resulted in one  $1000 \times 9$  matrix simulating test-level scores of 9 tests for 1000 subjects.

The psychometric network analyses were conducted with the “psychonetrics” package (v 0.9; Epskamp, 2021) in R. R Scripts for the simulation and the network analyses are available at <https://osf.io/syxed/>. Network models were estimated by modeling the variance-covariance matrix of the data as Gaussian graphical models (Epskamp et al., 2017). For the standard network model with simulated scores of the 9 tests as nodes, a model optimization procedure was conducted, in which the initial model was pruned by a step-down search process with a significance level of .01 in a recursive manner, such that edges that were not significant at  $\alpha = .01$  were automatically and recursively removed. The pruned model was then optimized by a step-up search process with a significance level of .01, such that the edges that were removed in the previous steps were added back, based on modification indices, until BIC no longer increased. For the latent network model, a total of 3 latent nodes were recognized based on the simulation specifications: Fluid, Verbal, and Spatial. Each latent node consisted of three corresponding tasks. Due to the small set of latent nodes, no model optimization process for the latent network was conducted. The latent network model was presented in its initial form of a graphical model.

In the current study, the higher-order factor model was also estimated for the simulated data for conceptual comparison of the latent factor model, network model, and latent network model. Specifically, to conceptually and statistically compare the broad ability factors and the network clusters, we estimated factor scores (and their equivalent statistics in network models) for the simulated broad abilities from the latent factor model, the network model, and the latent

network model. For the higher-order latent factor model, factor scores of the three broad ability factors (Gf, Gv, and Gs) were estimated by the regression method. Similarly, for the latent network model, due to the similarity between the latent nodes and the subfactors in the latent factor model, factor scores of the latent nodes (referred to as “latent node scores”) were estimated for the three latent nodes using the same regression method. For the standard network model, a network equivalent of factor score, which we referred to as the “cluster score”, was estimated (Golino et al., 2020). Previous research (Hallquist et al., 2019; Christensen & Golino, 2021) has demonstrated that one of the centrality measures of nodes in a network model, namely the node strength measure, represents a combination of dominant and cross-factor loadings for the node, as a manifest variable in a factor model. This can be mathematically represented as:

$$S_i = \sum_{f=1}^F L_{if} = \sum_{f=1}^F \sum_{j \in f} |\omega_{ij}| = \sum_{j=1}^n |\omega_{ij}| \quad \#(8)$$

Where  $S_i$  is the node strength of node  $i$  (the sum of all the absolute edge weights that directly connect  $i$ ),  $L_{if}$  is the “network loading” of  $i$  on the  $f$  cluster, which is the sum of the absolute edge weights in factor  $f$  that are directly connected to  $i$ .

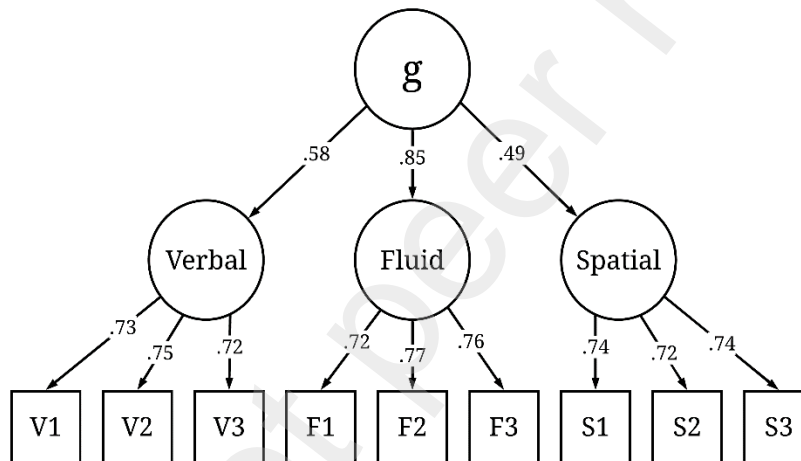
Previous studies have demonstrated that the network loadings can effectively estimate the effects from multiple latent causes on individual nodes and are comparable with exploratory or confirmatory factor loadings in the latent factor models (Christensen & Golino, 2021). Using these network loadings as corresponding weights of each item on a specific cluster of nodes, the network “cluster” scores can be estimated for each individual. Therefore, in the current study, we embraced this approach and estimated individual “cluster” scores for the three clusters of fluid tasks, verbal tasks, and spatial tasks, based on the result of the optimized standard network model.

## Study 2 Results

The higher-order latent factor model exhibited excellent fit to the simulated data:  $\chi^2(24) = 20.24, p = .683$ ; CFI = 1.00, TLI = 1.00; RMSEA < .01, SRMR = .02; AIC = 55450.35, BIC = 55553.41 (see Figure 3). The results of the simulation replicated the simulation in Study 1 (see Figure 3).

**Figure 3**

*The Higher-Order Latent Factor Model of the Simulated Data Based on POT*



*Note.* F1, F2, and F3 were fluid tasks; V1, V2, and V3 were verbal tasks; S1, S2, and S3 were spatial tasks. All factor loadings were statistically significant.

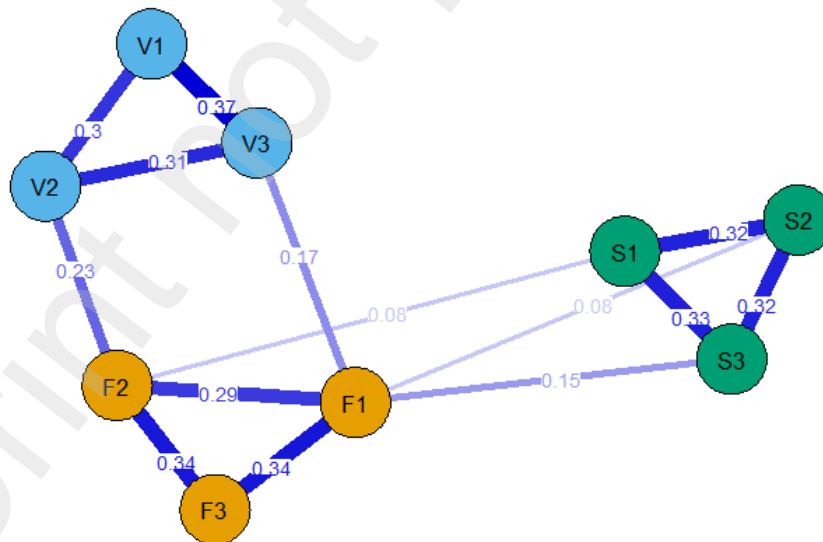
A standard network model analysis was applied to the full simulated dataset with nine manifest variables as nodes for the network. As described in the method section, the network model was pruned by a step-down search process and then optimized by a step-up search process. The standard network model exhibited great fit according to the fit statistics (with the exception of significant  $\chi^2$  statistics): for,  $\chi^2(22) = 98.93, p < .001$ , CFI = .97, TLI = .95, RMSEA = .06, AIC = 55804.93, BIC = 55961.98.



The weighted, undirected optimized standard network model of the nine tests simulated from the POT algorithm was presented in Figure 4. Nodes in this model, representing the manifest variables, were colored to reflect the theoretical clusters based on types of the corresponding tasks (Fluid, Verbal, and Spatial). Compared to the higher-order latent factor model, this standard network model described the same correlations among the nine simulated variables without assuming any latent factor. However, the three broad abilities were still represented by the three small clusters of nodes. Furthermore, compared to the verbal and spatial tests, the fluid tests played a more central role in the network model: the fluid nodes connected to both spatial and verbal nodes, while edges between the spatial and verbal nodes were not significant and therefore excluded. This pattern was a network representation of the larger factor loading of the fluid factor to the higher-order  $g$  in the latent factor model (Figure 3).

**Figure 4**

*The Standard Network Model of the Simulated Data Based on POT*



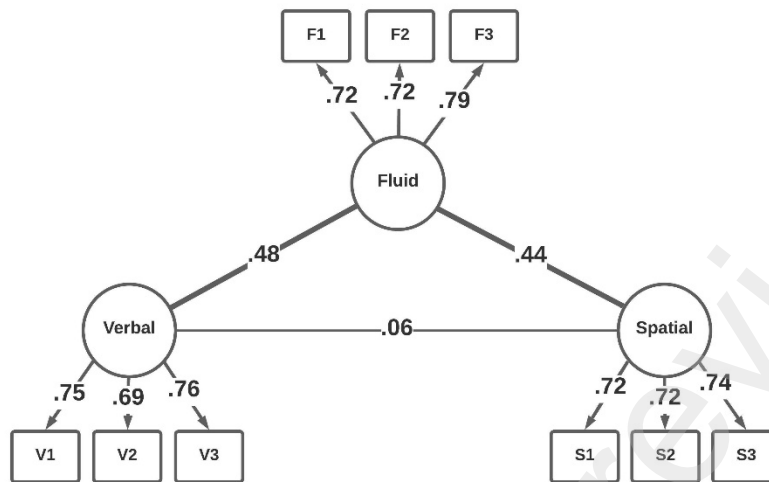
*Note.* The network was a weighted, undirected Gaussian graphical model optimized using step-down search and step-up search. F1, F2, and F3 were fluid tasks; V1, V2, and V3 were verbal tasks; S1, S2, and S3 were spatial tasks.

A latent network model analysis was conducted with the same dataset but with three latent variables (Fluid, Verbal, and Spatial) specified as nodes for the network, whereas the nine manifest variables loaded onto the three latent variables correspondingly. Due to the small set of latent nodes, no model optimization process for the latent network was conducted. The latent network model exhibited acceptable fit according to the fit statistics,  $\chi^2(24) = 179.19$ ,  $p < .001$ , CFI = .96, TLI = .93, RMSEA = .08, AIC = 22952.53, BIC = 23099.76. Figure 5 presented the result of the latent network model.

Unlike the former model, nodes in this model represented latent variables estimated from the nine manifest variables (presented as boxes in the figure). Similar to the former model, the conditional dependent associations (partial correlations) between these latent nodes were presented as weighted edges. Furthermore, factor loadings of the observed variables were presented as directed paths (arrows from latent nodes to manifest variables). In this latent network, the edges between Verbal and Fluid factors and Spatial and Fluid factors were statistically significant, both with  $p < .001$ . The edge between Verbal and Spatial was not statistically significant,  $p = .14$ . The latent network model, again, described the same correlations among the nine variables and combines features of the latent factor model and the network model. In the current latent network model, the broad ability nodes were positively associated with each other without attributing to a latent common cause. Furthermore, compared to either the Verbal or the Spatial node, the Fluid node was more central to the latent network.

### Figure 5

*The Latent Network Model of the Simulated Data*



*Note.* The latent network was a weighted undirected Gaussian graphical model without optimization. F1, F2, and F3 were fluid tasks; V1, V2, and V3 were verbal tasks; S1, S2, and S3 were spatial tasks. All factor loadings were statistically significant. The edge between Verbal and Spatial factors was not statistically significant ( $p = .14$ ).

As mentioned, for the fluid tasks, verbal tasks, and spatial tasks, factor scores were estimated for the latent factor model, cluster scores were estimated for the standard network model, and latent node scores were estimated for the latent network model. These estimated individual scores for the three broad abilities from the three types of models were compared. Results indicated that, for each of the three broad abilities, the estimated scores from the three models were almost perfectly correlated with each other (Table 2).

**Table 2.** *Correlations among factor and cluster scores for the fluid, verbal and spatial tasks.*

		Network Cluster Scores		
		Fluid	Verbal	Spatial
Latent Factor Scores	Fluid	<b>0.99</b>	0.62	0.61
	Verbal	0.62	<b>0.99</b>	0.36
	Spatial	0.61	0.36	<b>0.99</b>
		Latent Node Scores		
		Fluid	Verbal	Spatial
Latent Factor Scores	Fluid	<b>0.99</b>	0.60	0.58

	Verbal	0.59	<b>0.99</b>	0.33
	Spatial	0.58	0.33	<b>0.99</b>
<hr/>				
		Latent Node Scores		
		Fluid	Verbal	Spatial
	Fluid	<b>0.99</b>	0.62	0.61
Network Cluster Scores	Verbal	0.61	<b>0.99</b>	0.36
	Spatial	0.62	0.36	<b>0.99</b>

*Note.* Latent factor scores were estimated from the latent factor model. Network cluster scores were estimated from the standard network model. Latent node scores were estimated from the latent network model. The correlations of the estimated scores for the same broad ability construct in different models were all in **bold**.

## Discussion, Study 2

Study 2 introduced psychometric network models as an alternative approach for examining the correlational structure of cognitive abilities. These models estimate conditional dependent associations between test scores, effectively visualizing the interconnectedness of observed data without assuming unobserved latent factors (Schmank et al., 2019).

Compared to the conventional latent factor models, these network models are more compatible with the conceptual model of process overlap theory (POT-V), such that the correlations among cognitive testing data are caused by the overlaps among the sampling of cognitive processes involved in different tests. The network model represents these process overlaps as edges between nodes, with stronger edges suggesting a higher overlap of sampled cognitive processes between two tests. Thus, shared variance among variables is visualized as clusters of nodes with strong edges (Figure 4), rather than latent factors and factor loadings in a CFA model (Figure 3). In the POT algorithm, domain-general processes (defined as the executive function processes) were sampled across all 9 tests, and the three types of domain-specific processes (fluid, verbal, and spatial processes) were sampled only by their corresponding

types of tests. This sampling algorithm was represented by the current standard network model, as the 9 nodes were generally connected as a graphic network among the 9 nodes, and three broad abilities were represented by three clusters of corresponding nodes.

Notably, in the current study, fluid ability emerged as central in the network model results, with edges between fluid and other ability nodes being stronger than the edges between verbal and spatial nodes. This central role of fluid tests in the current network was also compatible with the common finding in psychometric models that the tests that are designed to directly measure fluid cognitive functions tend to have higher loadings on  $g$  than other types of tests (Gustafsson, 1988; Blair, 2006), often approximating statistical equivalence (i.e., a loading of 1). This result directly illustrated the POT algorithm's sampling mechanism, such that the EF processes were sampled with a higher probability ( $P = 0.28$ ) in fluid tests than in verbal and spatial tests ( $P = 0.12$ ), leading to greater overlaps of processes between fluid and other tests.

The latent network model of intelligence went a step further and combined the benefits of both latent factor models and psychometric network models. That is, the latent nodes of broad abilities (Fluid, Verbal, and Spatial in Figure 5) retained the benefits of latent factor models and accounted for measurement error in cognitive tasks; and the network structure among these latent nodes rejected the notion of general intelligence as a common cause of the positive manifold and reflected the connectedness of the broad ability factors using partial correlations. Similar to the edges in a standard network model, the partial correlation edges among latent nodes were similar to undirected alternatives to the directed regression paths among latent factors in structural equation models. Similar to the standard network model, the results of the latent network model also replicated the central role of fluid ability in the standard network model.

For the three broad ability constructs, the estimated factor/cluster scores from each of the three types of models were almost perfectly correlated. This result was aligned with previous findings in which the network “cluster” scores correlated with factor scores between .90 to .99 (Golino et al., 2020). Thus, the cluster scores and latent node scores in the standard and latent network model could also be used for obtaining information about individuals’ standing on the identified clusters/latent nodes. For instance, as part of a regression analysis, applied researchers could easily estimate the cluster score of verbal tasks from a network model as the network equivalent of a verbal factor score in a latent factor model. However, the interpretation of the estimated cluster scores may be different from factor scores. In fact, these cluster scores from the standard network models were conceptually more similar to formative latent variables than reflective latent variables (Christensen & Golino, 2021).

Overall, these results suggest that the latent factor model, which assumes a common cause, is not necessarily superior to or more informative than the other models. The network models, which are more compatible with POT, can capture the same relationships among intelligence test results in the absence of a “general intelligence”. Both standard and latent network models offer more flexibility and versatility compared to the latent factor model, with no or fewer assumptions of a psychological cause of the common variance.

### **General Discussion**

The current studies investigated a contemporary theoretical framework of intelligence, the process overlap theory (POT; Kovacs & Conway, 2016; 2019), that attempts to bridge the gap between psychometric and cognitive theories of human intelligence. POT proposed a novel psychometric structure and cognitive architecture to explain individual differences in cognitive abilities. Study 1 illustrated a simulation based on the sampling algorithm of POT (as was

proposed in POT-V and POT-I) to demonstrate that, under the POT algorithm, the positive manifold can emerge at the psychometric level in the absence of a general cognitive ability at the cognitive level. This simulation was developed to conceptually prove the sampling mechanism of POT at both the cognitive level and the psychometric level. In the POT algorithm, simulated cognitive processes were sampled and overlapped in different intelligence tests under regulations based on cognitive mechanisms. The psychometric results of the simulated data, in the form of latent factor models (POT-S), replicated the higher-order factor model observed in empirical psychometric theories such as the CHC theory. A standard higher-order “general intelligence” model fit the simulated testing data well, even though there was no general cognitive ability involved in generating the data.

Study 2 examined the simulation dynamics based on the POT algorithm in the form of network models (POT-N). This study aimed to investigate the correlational structure of the simulated cognitive testing scores based on POT using both a standard network model (Borsboom et al., 2021) and a latent network model (Epskamp et al., 2017). The standard network model is fully exploratory and data-driven, while the latent network model retains benefits from both exploratory and confirmatory approaches. In study 2, the standard network model presented a network of the nine simulated test scores that revealed the three broad abilities. The latent network model replicated similar findings from the standard network model, but under confirmatory assumptions regarding the broad abilities as latent factors (latent nodes). In both models, measures of fluid ability played a more central role than measures of either spatial or verbal ability. This emphasis on fluid ability in the two network models reflected the sampling mechanism proposed by POT, in which domain-general cognitive processes were sampled more often than domain-specific processes across different types of tests. The latent

network model was more compatible with POT-V and POT-S than the standard network model, in which the latent nodes are equivalent to the latent broad ability factors. The formative  $g$  in POT-S is, in fact, conceptually comparable to the network structure of the latent nodes: both of them are determined by the indicators rather than being the cause of the variance among indicators.

Overall, these current results demonstrated that, based on the POT algorithm, the positive manifold of intelligence can emerge at the psychometric level in the absence of a general mental ability at the cognitive level. Moreover, a reflective model that assumes a causal general factor fits the data well where in fact no general causal mechanism had any role in generating the data. This study provides critical supportive evidence for POT and illustrates an alternative theoretical and statistical framework for contemporary research of human cognition that combines psychometric and cognitive theories of intelligence.

There are, of course, potential limitations of the current studies that could inspire corresponding future research. An important limitation of the current POT simulation framework is that in the current sampling mechanism, the “cognitive processes” are only random numbers that are conceptually equivalent. The cognitive distinctions are specified only by the sampling mechanism such as the sampling probabilities. This could be updated by introducing different types of cognitive processes in the simulation that could account for more behavioral and neural science observations, such as simulated cognitive processes/mechanisms that could reflect the statistical parameters and cognitive properties of processing speed. One potential direction is to integrate the drift-diffusion model (Molenaar et al., 2015) in the current POT-I algorithm. The drift-diffusion model, combined with the current POT-I simulation framework, may represent a



more realistic decision-choice mechanism that is related to response time parameters in cognitive tasks.

Another important limitation of the current simulation framework is that the simulated tests are relatively simple compared to empirical observations in cognitive and psychometric research. In the current simulation framework, there are a total of 4 types of cognitive processes and each of the tasks only samples no more than 2 types. This is highly unlikely in the real world. More types of simulated cognitive processes and complex tasks could be introduced in the current simulation method.

For the network modeling aspect, the current study only applied the psychometric network models to a simulated dataset of cognitive tests. Given the exploratory nature of network psychometrics, it is important to apply this contemporary psychometric approach to empirical samples before reaching a generalizable interpretation of the domain-general and domain-specific emphasis based on POT. Network psychometrics is also developing rapidly, and recent techniques encourage more confirmatory analyses using network models such as multi-group invariance comparison, time-series modeling, or even meta-network analysis (Isvoranu et al., 2021). Network models using these techniques could be more comparable to conventional psychometric models of behavioral data as well as neuroscience data. The applications of network psychometrics could inspire unification between psychometric and cognitive research and improve the measurement, explanation, and prediction of individual differences in human intellectual behaviors at different levels of analysis.

## References

- 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), 567-579. <https://doi.org/10.1037/a0016262>
- Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences*, 29(2), 109-125. doi:10.1017/S0140525X06009034
- Boring, E. G. (1923). Intelligence as the tests test it. *New Republic*, 35(6), 35-37.
- Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., ... & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, 1(1), 1-18. <https://doi.org/10.1038/s43586-021-00055-w>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203–219. <https://doi.org/10.1037/0033-295X.110.2.203>
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies* (No. 1). Cambridge University Press.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1–22. <https://doi.org/10.1037/h0046743>

- Christensen, A. P., & Golino, H. (2021). On the equivalency of factor and network loadings. *Behavior Research Methods*, 53(4), 1563-1580. <https://doi.org/10.3758/s13428-020-01500-6>
- Conway, A. R. A, & Kovacs, K. (2015). New and emerging models of human intelligence. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(5), 419-426. <https://doi.org/10.1002/wcs.1356>
- Conway, A. R. A, Kovacs, K., Hao, H., Goring, S. A., & Schmank, C. (2020). The Struggle is Real: Challenges and Solutions in Theory Building. *Psychological Inquiry*, 31(4), 302-309. DOI: [10.1080/1047840X.2020.1853468](https://doi.org/10.1080/1047840X.2020.1853468)
- Conway, A. R. A., Kovacs, K., Hao, H., Rosales, K. P., & Snijder, J.-P. (2021). Individual Differences in Attention and Intelligence: A United Cognitive/Psychometric Approach. *Journal of Intelligence*, 9(3), 34. <https://doi.org/10.3390/jintelligence9030034>
- Deary, I. J. (2003). Reaction time and psychometric intelligence: Jensen's contributions. In *The Scientific Study of General Intelligence* (pp. 53-75). Pergamon. <https://doi.org/10.1016/B978-008043793-4/50040-4>
- Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, 11(3), 201-211. <https://doi.org/10.1038/nrn2793>
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19-23. DOI: 10.1111/1467-8721.00160

Epskamp, S (2021). Psychonetrics: Structural Equation Modeling and Confirmatory Network Analysis. R package version 0.9. <https://CRAN.R-project.org/package=psychonetrics>

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, 50(1), 195-212.  
<https://doi.org/10.3758/s13428-017-0862-1>

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, 23(4), 617. <https://doi.org/10.1037/met0000167>

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, 82(4), 904-927.  
<https://doi.org/10.1007/s11336-017-9557-x>

Eysenck, H. J. (1987). Thomson's "Bonds" or Spearman's "Energy": Sixty Years On. *Mankind Quarterly*, 27(3), 259.

Flanagan, D. P., Alfonso, V. C., & Reynolds, M. R. (2013). Broad and narrow CHC abilities measured and not measured by the Wechsler scales: Moving beyond within-battery factor analysis. *Journal of Psychoeducational Assessment*, 31(2), 202-223.  
<https://doi.org/10.1177/0734282913478047>

Frischkorn, G. T., & Schubert, A. L. (2018). Cognitive models in intelligence research: Advantages and recommendations for their application. *Journal of Intelligence*, 6(3), 34.  
<https://doi.org/10.3390/jintelligence6030034> Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., ... & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the

- number of latent factors: A simulation and tutorial. *Psychological Methods*, 25(3), 292–320. <https://doi.org/10.1037/met0000255>
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24(1), 79-132. [https://doi.org/10.1016/S0160-2896\(97\)90014-3](https://doi.org/10.1016/S0160-2896(97)90014-3)
- Gustafsson, J. E. (2014). Hierarchical models of individual differences in cognitive abilities. In *Advances in the psychology of human intelligence* (pp. 35-71). Psychology Press.
- Hallquist, M. N., Wright, A. G., & Molenaar, P. C. (2021). Problems with centrality measures in psychopathology symptom networks: Why network psychometrics cannot escape psychometric theory. *Multivariate behavioral research*, 56(2), 199-223. <https://doi.org/10.1080/00273171.2019.1640103>
- Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, 75(3), 242–259. <https://doi.org/10.1037/h0025662>
- Hunt, E., Lunneborg, C., & Lewis, J. (1975). What does it mean to be high verbal? *Cognitive Psychology*, 7(2), 194–227. [https://doi.org/10.1016/0010-0285\(75\)90010-9](https://doi.org/10.1016/0010-0285(75)90010-9)
- Isvoranu, A. M., Epskamp, S., & Cheung, M. W. L. (2021). Network models of posttraumatic stress disorder: A meta-analysis. *Journal of Abnormal Psychology*, 130(8), 841. <https://doi.org/10.1037/abn0000704>
- Jensen, A. R. (1998). The suppressed relationship between IQ and the reaction time slope parameter of the Hick function. *Intelligence*, 26(1), 43-52. [https://doi.org/10.1016/S0160-2896\(99\)80051-8](https://doi.org/10.1016/S0160-2896(99)80051-8)

- Kan, K. J., van der Maas, H. L., & Levine, S. Z. (2019). Extending psychometric network analysis: Empirical evidence against g in favor of mutualism?. *Intelligence*, 73, 52-62. <https://doi.org/10.1016/j.intell.2018.12.004>
- Kaufman, J. C., Kaufman, S. B., & Plucker, J. A. (2013) Contemporary Theories of Intelligence. In *The Oxford Handbook of Cognitive Psychology*.
- Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, 27(3), 151-177. <https://doi.org/10.1080/1047840X.2016.1153946>
- Kovacs, K., & Conway, A. R. A. (2019). A unified cognitive/differential approach to human intelligence: Implications for IQ testing. *Journal of Applied Research in Memory and Cognition*, 8(3), 255-272. <https://doi.org/10.1016/j.jarmac.2019.05.003>
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence*, 14(4), 389-433. [https://doi.org/10.1016/S0160-2896\(05\)80012-1](https://doi.org/10.1016/S0160-2896(05)80012-1)
- McGrew, K. S. (2005). The Cattell-Horn-Carroll Theory of Cognitive Abilities: Past, Present, and Future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 136–181). The Guilford Press.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 1(37), 1-10. <https://doi.org/10.1016/j.intell.2008.08.004>

- Molenaar, D., Tuerlinckx, F., & van der Maas, H. L. (2015). Fitting diffusion item response theory models for responses and response times using the R package diffIRT. *Journal of Statistical Software*, 66(4), 1-34. <https://doi.org/10.18637/jss.v066.i04>
- Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., ... & Urbina, S. (1996). Intelligence: knowns and unknowns. *American Psychologist*, 51(2), 77. <https://doi.org/10.1037/0003-066X.51.2.77>
- Roid, G. H., & Pomplun, M. (2012). *The Stanford-Binet intelligence scales*. The Guilford Press.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Savi, A. O., Marsman, M., van der Maas, H. L., & Maris, G. K. (2019). The wiring of intelligence. *Perspectives on Psychological Science*, 14(6), 1034-1061. <https://doi.org/10.1177/1745691619866447>
- Schmank, C. J., Goring, S. A., Kovacs, K., & Conway, A. R. (2019). Psychometric network analysis of the Hungarian WAIS. *Journal of Intelligence*, 7(3), 21. <https://doi.org/10.3390/jintelligence7030021>
- Schrank, F. A. (2011). Woodcock-Johnson III Tests of Cognitive Abilities. In A. S. Davis (Ed.), *Handbook of pediatric neuropsychology* (pp. 415–434). Springer Publishing Company.
- Schrank, F. A., Decker, S. L., & Garruto, J. M. (2016). *Essentials of WJ IV cognitive abilities assessment*. John Wiley & Sons.

- 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*. London: Macmillan.
- Sternberg, R. J. (1985). Human intelligence: The model is the message. *Science*, 230(4730), 1111-1118. <https://doi.org/10.1126/science.230.4730.1111>
- Sternberg, R. J. (2013). Contemporary theories of intelligence. In W. M. Reynolds, G. E. Miller, & I. B. Weiner (Eds.), *Handbook of psychology: Educational psychology* (pp. 23–44). John Wiley & Sons, Inc.
- Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, 1904-1920, 8(3), 271–281. <https://doi.org/10.1111/j.2044-8295.1916.tb00133.x>
- Thomson, G. (1951). *The factorial analysis of human ability*. University of London Press.
- Thorndike, E. L. (1925). *The measurement of intelligence*. New York: Columbia University, Teachers College.
- Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113(4), 842–861. <https://doi.org/10.1037/0033-295X.113.4.842>
- Van Der Maas, H. L., Kan, K. J., Marsman, M., & Stevenson, C. E. (2017). Network models for cognitive development and intelligence. *Journal of Intelligence*, 5(2), 16. <https://doi.org/10.3390/jintelligence5020016>