

The Version of Scholarly Record of this Article is published in *Research in Developmental Disabilities*, available online at: [http://dx.doi.org/ 10.1016/j.intell.2019.101380](http://dx.doi.org/10.1016/j.intell.2019.101380). Note that this article may not exactly replicate the final version published in *Intelligence*.

Giofrè, D., Pastore, M., Cornoldi, C., & Toffalini, E. (2019). Lumpers vs. splitters: Intelligence in children with specific learning disorders. *Intelligence*, 76, 101380.  
<https://doi.org/10.1016/j.intell.2019.101380>

### **Lumpers vs. splitters: Intelligence in children with specific learning disorders**

David Giofrè<sup>1</sup>, Massimiliano Pastore<sup>2</sup>, Cesare Cornoldi<sup>3</sup>, Enrico Toffalini<sup>3</sup>

<sup>1</sup> School of Natural Science and Psychology, Liverpool John Moores University, UK

<sup>2</sup> Department of Developmental and Social Psychology, University of Padova, Italy

<sup>3</sup> Department of General Psychology, University of Padova, Italy

Correspondence to be addressed to:

Dr. David Giofrè,  
School of Natural Sciences and Psychology,  
Liverpool John Moores University, UK  
2.28a James Parsons Building,  
Byrom Street, Liverpool, L3 3AF

**Acknowledgements:** We would like to thank David Recine and Ingrid Boedker for the editing and helpful comments.

### **Abstract**

A multigroup bifactor model was used to compare the explained variance and the reliability of the general vs. specific factors of the Wechsler Intelligence Scale for Children-4th edition (WISC-IV) in 1617 Italian children diagnosed with specific learning disorder (SLD) and an Italian normative sample of typically-developing children (with the exclusion of  $IQ < 70$ ). Results suggested that more than half of the common variance, 56.1%, was accounted for by the domain-specific factors in SLD, against only 39.5% in typical development. The reliability of both general and specific factors was rather limited in SLD, whereas the reliability of the g-factor was good in typical development. An additional analysis using previous information from American data showed very similar results. Our results suggest that the role of the specific factors, of VCI and PSI in particular and WMI to a lesser extent, should be considered as probably largely distinct from the g-factor in children with SLD. Results also seem to indicate that the PRI is a less distinctive factor, which is, in the SLD group, hardly distinguishable from the g-factor. The use of Bayesian priors from American data indicated that results on Italian and American samples of children with SLD were similar, and different from those on the normative samples of both countries, suggesting remarkable cross-cultural and cross-linguistic similarity of the structure of intelligence in children with SLD.

### **Lumpers vs. splitters: Intelligence in children with specific learning disorders**

Intelligence involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience ([Gottfredson, 1997](#)). A harshly debated question in intelligence research has been whether one's "intelligence" is better described by a single overarching factor, or by a set of several separate skills ([Willis, Dumont, & Kaufman, 2011](#)). On one hand, [Spearman \(1904\)](#) emphasized the interpretation of intelligence as a single factor (or *g*-factor). At the other extreme, some authors have emphasized the importance of a set of distinct abilities within human intelligence (e.g., [Guilford, 1967](#); [Horn & Cattell, 1966](#); [Thorndike, 1903](#)). These two different views have been described as "lumpers" who hold that intelligence can be largely subsumed by a single factor, and "splitters" who hold the view that intelligence is better described by several distinct factors ([Willis et al., 2011](#)). With the introduction of hierarchical theories of intelligence, these two opposing approaches have been integrated into a single framework, assuming that intelligence can be fully explained by the combination of specific and general aspects ([Carroll, 1993](#); [Vernon, 1950](#); see also [Johnson & Bouchard, 2005](#)).

Several batteries exist for the assessment of intelligence, but the Wechsler Intelligence Scale for Children, 4th edition (WISC-IV; [Wechsler, 2003](#)), is arguably the most popular instrument for the assessment of children throughout the world ([Evers et al., 2012](#)). The WISC-IV allows for both a single intelligence index (FSIQ) and four primary indices, including Perceptual Reasoning (PRI), Verbal Comprehension (VCI), Working Memory (WMI), and Processing Speed (PSI). The presence of both a single *g*-factor index (i.e., FSIQ), as well as other indices, makes it possible to evaluate both a child's general intellectual ability and their cognitive profile, obtained from different indices. For this reason, the use of the WISC-IV appears to be particularly appropriate for testing the existence and the clinical utility of, respectively, the single *g*-factor and the specific indices.

There is an extensive literature on the utility of principal indices vs. a global index in both clinical and non-clinical samples. On one hand, some authors (i.e., "lumpers") hold the view that

only the FSIQ has practical utility, as it is more reliable and informative compared to the other principal indices. This position is supported by evidence indicating that the FSIQ is in fact a robust predictor of achievement, which explains the large majority of the subtests variance, at least in the general population, leaving little margin for other specific factors (e.g., [Watkins, Glutting, & Lei, 2007](#)). This is also supported by studies indicating that the FSIQ tend to be more reliable than the other indices (e.g., [Watkins, 2006](#); [Watkins & Smith, 2013](#)). Conversely, other scholars (i.e., splitters) caution that the FSIQ may distort the interpretation of a child's intellectual functioning, at least when principal indices are scattered and the profile is not flat (e.g., [Cornoldi, Giofrè, Orsini, & Pezzuti, 2014](#); [Flanagan & Kaufman, 2004](#); [Fiorello, Hale, Mcgrath, Ryan, & Quinn, 2001](#), [Fiorello et al., 2007](#); [Sattler & Saklofske, 2001](#); [Toffalini, Giofrè, & Cornoldi, 2017a](#)). These scholars also tend to believe that the principal WISC-IV indices have a sufficient explanatory power when used in association with the FSIQ (e.g., [Fiorello et al., 2007](#); [Giofrè, Toffalini, Altoè, & Cornoldi, 2017](#); [Toffalini, Giofrè, & Cornoldi, 2017b](#)). Support for this view is particularly evident from the study of children with atypical development.

There is a large amount of research indicating that the cognitive profile of children with various neurodevelopmental disorders tends to be uneven, as it happens in children with specific learning disorders (SLD, [Cornoldi et al., 2014](#); [De Clercq-Quaegebeur et al., 2010](#); [Poletti, 2016](#); [Styck & Watkins, 2016](#)), with autism spectrum disorder ([Giofrè et al., 2019](#); [Mayes & Calhoun, 2008](#); [Oliveras-Rentas, Kenworthy, Roberson, Martin, & Wallace, 2012](#)), ADHD ([Mayes & Calhoun, 2007](#); [Thaler, Bello, & Etcoff, 2013](#)), and developmental coordination disorder ([Loh, Piek, & Barrett, 2011](#); [Sumner, Pratt, & Hill, 2016](#)). Also for this reason, the majority of practitioners who work with children in clinical settings continue to rely on the WISC-IV indices rather than the FSIQ alone ([Flanagan & Kaufman, 2004](#); [Pfeiffer, Reddy, Kletzel, Schmelzer, & Boyer, 2000](#)).

Extensive research on the structure of the WISC-IV is available pertaining to both children with typical and atypical development. Most of the study of the structure of the WISC-IV examines

the higher order relationship of the g-factor with first-order factors, within the so-called hierarchical model (see [Canivez, 2014](#) on this point). Alternative models, including the so-called bifactor model have been proposed (see [Gignac, 2016](#), for a review and [Molenaar, 2016](#) for a different qualification).

The bifactor model specifies that the g-factor has a direct effect on the subtests, whereas the specific factors (VC, PR, WM, and PS) predict the correlated residual variance of the subtests. When this model is adopted, the g-factor and the specific indices are usually specified to be orthogonal. As a result, the specific factors capture the domain-specific variability, with the effects of the g-factor and the other indices partialled out. The bifactor model has been repeatedly used with WISC-IV data across several different cultures and samples (e.g., [Canivez, 2014](#); [Canivez, Watkins, Good, James, & James, 2017](#); [Gignac & Watkins, 2013](#); [Golay, Reverte, Rossier, Favez, & Lecerf, 2013](#); [McGill & Canivez, 2017](#); [Styck & Watkins, 2016](#); [Watkins, Canivez, James, James, & Good, 2013](#)). This model is generally adopted to establish the proportion of subtest variance attributed to general vs. specific factors, as well as the reliability of general and specific factor scores.

Other approaches to the same research questions can be found in the literature. Namely, an exploratory factor analysis based on the Schmid-Leiman transformation on the correlation matrix and a predefined number of factors has been suggested ([Schmid & Leiman, 1957](#)). This approach is based on an unrestricted solution and does not require an a priori underlying theoretical model. Alternatively, a confirmatory solution can be used by applying a Schmid-Leiman decomposition of the coefficients of a (restricted) higher-order model solution (e.g., [Gignac & Kretzschmar, 2017](#)), which is the most frequently adopted model in the research on the Wechsler scales. Such a decomposition is equivalent to calculating the indirect effects of the g-factor on subtests in a model in which the first-order factors act as mediators.

The higher-order and the bifactor models are not equivalent. In a higher-order model the effect of the g-factor on a specific subtest is necessarily mediated by a first-order factor. Therefore,

the relation between the *g*-factor and a specific subtest is forced to be proportional to the relation between the *g*-factor and the corresponding first-order factor. This so-called “proportionality constraint” has already been described in the literature, and is not imposed when using a bifactor model (see [Gignac, 2016](#) for further details, and [Mansolf & Reise, 2017](#), for a comprehensive discussion).

In the present study, we examined the results obtained with a bifactor model solution, in comparison to a Schmid-Leiman decomposition of the higher-order model. Confirmatory (i.e., restricted) solutions were used for both models. We used data collected in two different countries, namely Italy and the USA, on children with typical development and with Specific Learning Disabilities (SLD) tested using the WISC-IV. The WISC-IV has been validated and evaluated across different languages and cultures, generally pointing to large similarities across different contexts (e.g., [Chen, Keith, Weiss, Zhu, & Li, 2010](#); [Reverte, Golay, Favez, Rossier, & Lecerf, 2015](#)). Some evidence has already been collected assessing clinical groups, including several types of different disorders and subclinical conditions (e.g., [Canivez, 2014](#); [Chen & Zhu, 2012](#)). However, whether the structure of the WISC-IV is invariant across normative and clinical samples is not fully understood.

It has been recently suggested that there is configural invariance between specific learning disorder (SLD) and typical development in the Italian population, but not metric invariance ([Giofrè & Cornoldi, 2015](#)). In [Giofrè and Cornoldi \(2015\)](#), the results showed that the PRI and VCI are strongly associated with the *g*-factor in both SLD and the typically-developing population, whereas the other indices (e.g., WMI, PSI) are more weakly related to the *g*-factor, and this is particularly true in children with SLD as compared to the typically-developing children. Based on these premises, the authors recommended the use of visuo-perceptual subtests, or the use of the General Ability Index (GAI, a measure that combines the VCI and PRI subtests, but excludes WMI and PSI) as a reliable measure of the intellectual functioning of children with SLD. The authors, however, did not focus on the utility of the specific indices per se, because their study only focused on the

identification of reliable indices of intellectual functioning in children with SLD, and on testing various forms of invariance between the two groups. For this reason, the interpretability and reliability of specific indices, and the amount of variance explained for by each factor, was overlooked. Importantly, the aforementioned paper was focused on children from the Italian population and could hardly be generalized to different languages and contexts. This is particularly relevant for children with SLD, who typically present with difficulties in reading and writing, i.e., in abilities concerning the use of language, and who are therefore affected by the specific characteristics of specific languages. On that note, it is worth mentioning that Italian is a very transparent language, where there is consistent correspondence between graphemes and phonemes, while in other languages, such as English, graphemes do not consistently map to phonemes (see [Provazza, Giofrè, Adams, & Roberts, 2019](#) for further details).

For English speakers, different factorial models were applied to the American data from a large sample of children with SLD ([Styck & Watkins, 2016](#)). In their conclusions, the authors highlighted only the importance of considering FSIQ rather than the specific factors. A direct comparison between clinical and non-clinical groups, however, was not made. With regard to the American typically-developing population, a similar, albeit not identical, analytical approach had been reported by [Watkins \(2006\)](#).

With the present study, our first goal was to quantify the amount of the WISC-IV common variance captured by the g-factor alone in comparison to the other specific indices, in a large clinical group of Italian children diagnosed with SLD. We also aimed to quantify the reliability of both the specific indices and the global composite score. To address these research questions, we used a multigroup bifactor model approach for the first time, because it has the unique advantage of allowing us to establish the role of the domain-specific factors after the g-factor has been removed ([Gignac & Watkins, 2013](#)).

Children with neurodevelopmental disorders, and SLD in particular, represent an interesting case because they seem to be characterized by specific neuropsychological deficits that affect some

specific abilities, while leaving others relatively preserved (e.g., [D'Angiulli & Siegel, 2003](#); [Peng & Fuchs, 2016](#)). Scattered profiles have often been reported in these populations (e.g., [De Clercq-Quaegebeur et al., 2010](#); [Poletti, 2016](#); [Toffalini et al., 2017a](#), [Toffalini et al., 2017b](#)). With this in mind, it can be hypothesized that in children with SLD, scores on the specific factors might reflect the functioning in areas that have been spared (PRI), partially spared (VCI), or compromised to a various degree (WMI, PSI) by the learning disorder. It can also be hypothesized that the g-factor would probably tend to be stronger in children with typical development, as compared to children with SLD.

An additional goal of the present study was to compare the Italian and the American populations of children with SLD. This goal was achieved by using previously published information from the American population as a set of informative priors in Bayesian structural equation modelling (BSEM). This analysis was aimed to verify whether data collected with Italian children could be generalized to other countries with opaque orthographies. In other words, we aimed to ascertain whether the differences between SLD and the typically-developing population point in the same direction in contexts as different as Italy and the US. This is relevant to the more general question of whether the comparison between typical and atypical development can be generalized across cultures and languages. Notably, the Bayesian approach allows us to overcome some limitations of traditional approaches, and to directly apply previous information as informed priors in the statistical analysis (see [Golay et al., 2013](#) on this point).

## **1. Method**

### **1.1. Participants**

A sample of 1617 children ( $M_{\text{age}} = 11.57$  years, [ $SD = 2.45$ ]; 39.1% females), aged between 7 and 16 and with a diagnosis of SLD, was examined. WISC-IV profiles were collected under the auspices of the Italian Association for Learning Disabilities (AIRIPA), and were provided by 27 expert psychologists specialized in the assessment and treatment of SLD, working at clinical centers



in eight major Italian regions. Following the guidelines published by the National Italian Consensus Conference on Learning Disability ([Istituto Superiore di Sanità, 2011](#)), all diagnosed children met the following criteria: i) academic achievement, as assessed using standardized tests, was below the 5th percentile or 2 SDs below average in at least one specific area of learning; and ii) the low academic achievement could not be attributed to socio-cultural or educational deprivation, nor to sensory, neurological, or intellectual deficit; all children had an overall full scale IQ  $\geq 70$ . Notably, in Italy the WISC-IV is used to exclude the presence of an intellectual disability, but not to detect symptoms of SLD, nor to obtain a quantitative measure of the discrepancy between intelligence and learning. According to the Italian national guidelines, the ICD-10 coding system ([World Health Organization, 1993](#)) was used, in which the SLD corresponds to the F81 category (specific developmental disorder of scholastic skills).<sup>1</sup>

Previously published articles examined subsets of the same data ([Giofrè et al., 2017](#); [Giofrè, Stoppa, Ferioli, Pezzuti, & Cornoldi, 2016](#); [Toffalini et al., 2017a](#); [Toffalini, Pezzuti & Cornoldi, 2017](#)), but, focused on different aspects. None of the analyses presented in this study have been previously reported, have used the bifactor model, or have compared the Italian and American samples.

## **1.2. Instrument**

The Italian standardization of the WISC-IV (Orsini, Pezzuti, & Picone, 2012) was used. The scaled scores on the 10 basic subtests were considered.

## **1.3. Statistical approach**

A Bayesian approach to data analysis was adopted. The general advantages of the use of a Bayesian approach in psychological science have already been described in the literature (e.g., [Kruschke & Liddell, 2015](#); [van de Shoot et al., 2014](#); [Vandekerckhove, Rouder, & Kruschke, 2018](#)). In the present case, the main advantage is the ability to quantify evidence with precision when comparing the two groups through the multi-group analysis, particularly in terms of the overlap between posterior distribution ([Pastore, 2018](#)) of the estimated parameters in the two

groups, and the evidence ratios. This is more informative than the classic series of simplified accept-reject decisions of the null hypothesis significance testing (NHST) based on the  $p$ -value ([Kruschke & Liddell, 2015](#)). Another advantage of the Bayesian approach is that it allows for the application of prior knowledge directly into the analysis. This was crucial in a second phase of the analysis, in which prior evidence from American data was formalized as a series of informative priors.

With regard to the SLD sample, the scaled scores obtained by the 1617 diagnosed children in the 10 basic WISC-IV subtests were entered in the analysis. Given that the Bayesian MCMC method estimates the models directly from the likelihood function of the data, rather than from their correlation/covariance matrices, a simulation approach was used with regard to the typically-developing population. The typically-developing group was thus simulated from the correlation matrix contained within the Italian manual of the WISC-IV ([Orsini et al., 2012](#)), which was calculated based on the standardization sample of 2200 children between 6 and 16 years of age, under the assumption of normality of the distributions (as is appropriate in intelligence tests). To simulate a typically-developing population that is realistically comparable with the SLD sample, only simulated cases with FSIQ  $\geq 70$  were considered in the analysis.<sup>2</sup> Eventually, only 1617 simulated profiles (instead of 2200) were used for the typically-developing group in order to match the  $N$  of the SLD sample. For more details, see [Giofrè et al. \(2017\)](#), in which a very similar simulation procedure was followed.

A bifactor model was applied simultaneously to the SLD, and to the typically-developing group. In the bifactor model, the observed variables (i.e., the 10 basic subtests) have direct loadings on the  $g$ -factor, and four specific factors corresponding to VC, PR, WM, and PS. The bifactor model is orthogonal by definition, i.e., inter-correlations across all latent factors, constrained to be zero. Moreover, loadings of subtests on their specific factors were constrained to be equal. This solution was selected to ensure that the residual variance of different subtests (i.e., the variance after partialling-out the  $g$ -factor) equally contributes to their specific factor. This approach is warranted

to avoid major unbalances in the role of each subtest, in light of the limited number of subtests for each specific factor in the WISC-IV; furthermore, this solution allowed an easier model convergence (see [Beaujean, 2014](#) for a similar approach). Finally, the traditional four-factor higher-order model, with a superordinate g-factor ([Flanagan & Kaufman, 2004](#)), was also applied for a comparison with the bifactor solution.

All analyses were performed using the R software, version 3.5.0 ([R Core Team, 2018](#)) and Bayesian models were fitted using the “blavaan” package ([Merkle & Rosseel, 2018](#)), which is a package for estimating Bayesian structural equation models via JAGS ([Plummer, 2003](#)). In the main analysis, a set of non-informative priors was used, i.e., normal distributions with mean = 0, and precision = 0.01 (SD = 100) for all loading coefficients. In Bayesian terms, the priors are a set of information about the model parameters available before the data are observed, and they may have a leverage on the final estimates when set as informative. The loading of the first subtest in each factor was truncated above zero to obtain loadings with positive values (this is warranted because all subtests must have non-negative associations with any factor of intelligence). Posterior distributions for each parameter were estimated using four Markov chain Monte Carlo (MCMC), each running for 20,000 replicates. In Bayesian terms, the posterior distributions are probability density functions of the models coefficients. The final posterior estimates were calculated as the mean values of the posterior distributions. Highest Posterior Density intervals (HPDI; these are somehow comparable to the “confidence interval” in the frequentist framework) were reported as measures of uncertainty. All reported coefficients were standardized. Convergence of the model parameters was assessed by calculating the potential scale reduction factor (PSRF; [Gelman & Rubin, 1992](#)); this statistic measures the ratio of the average variance of samples within each chain to the variance of the pooled samples across chains; if all chains are at equilibrium, these will be the same and PSRF will be one.

An additional analysis was conducted using a set of informed priors formalized on the available American data. With regard to the typically-developing sample, the set of priors was

calculated using the maximum likelihood estimation method from the correlation matrix of the subtests presented in the American standardization of the WISC-IV ([Wechsler, 2003](#)). The number of observations was fixed to 1617 for consistency with the *Ns* of the present study. The estimated coefficients were used as the mean prior values, while precision was set to 1000.00 (i.e., standard deviation of the prior distribution was set to 0.03), which approximately reflects the standard errors of the estimates. Therefore, these can be considered as highly informed priors. With regard to the SLD sample, the estimated bifactor model coefficients reported by [Styck and Watkins \(2016\)](#) were used as the mean prior values. As measures of uncertainty for their estimated values were not reported in the original study, the same precision used for the typically-developing sample was employed. The full set of the mean values used for the informed priors is reported in the Supplemental online material. Thanks to the informed priors, no equality constraints were needed.

With regard to the model fit indices, the Bayesian root mean square approximation (BRMSEA;  $\geq 0$ , small is good), the Bayesian standardized root mean square residual (BSRMR;  $\geq 0$ , small is good), the Bayesian comparative fit index (BCFI; [0,1], large is good), and the coefficient of determination (CD; [0,1], large is good) were used. Between-model comparisons were based on the widely applicable information criterion (WAIC; [Watanabe, 2013](#); better models have lower values), and the Bayes Factor. Evidence ratios (ER) were calculated in terms of the proportion between the number of the posterior samples in which the SLD group had a coefficient greater (or smaller, depending on the type of comparison) than the typically-developing group, and the number of posterior samples in which the opposite happened. The reliabilities were calculated from the model coefficients using the  $\omega_h$  formula for the global composite score, and the  $\omega_s$  formula for the domain-specific indices ([Gignac & Watkins, 2013](#); [Reise, Bonifay, & Haviland, 2013](#)).

## 2. Results

### 2.1. Main analysis

Means and SDs of the WISC-IV subtests and factors in the two groups can be found in the Supplemental online material, Table S1. A preliminary analysis was conducted to compare the bifactor model with the higher-order model and the one-factor model. The results suggested that both the bifactor and the higher-order solutions had a very close fit with the data, with  $BRMSEA = 0.04$ , and  $BCFI = 0.97$ , in the bifactor model, and  $BRMSEA = 0.05$ , and  $BCFI = 0.97$ , in the higher-order model. However, the WAIC suggested that the bifactor model had a better fit than the higher-order model,  $\Delta WAIC = 10.70$ . The Bayes Factor (BF) also supported the bifactor model as more evident than the higher-order model,  $\log BF = 20.16$  (with reference to [Raftery, 1995](#), this could be considered as “strong” evidence in favor of the bifactor model). The bifactor model was also clearly better than the one-factor model (i.e., the one having the g-factor only),  $\Delta WAIC = 1445.03$  (in addition, the one-factor model had insufficient fit,  $BRMSEA = 0.12$ ,  $BCFI = 0.77$ ).

The fit indices of the bifactor models were very good both for the SLD group,  $BRMSEA = 0.04$ , 90% HPDI (0.04, 0.05),  $SRMR = 0.04$ , 90% HPDI (0.03, -0.04),  $BCFI = 0.97$ , 90% HPDI (0.97, 0.98),  $CD = 0.98$ , 90% HPDI (0.98, -0.99), and the typically-developing group,  $BRMSEA = 0.04$ , 90% HPDI (0.04, 0.05),  $SRMR = 0.02$ , 90% HPDI (0.02, 0.03),  $BCFI = 0.98$ , 90% HPDI (0.97, 0.98),  $CD = 0.92$ , 90% HPDI (0.91–0.94). Importantly, the multi-group model was better than the one with loadings fixed between-group,  $\Delta WAIC = 6.00$ , indicating that the two groups have overall different loadings.

It is worth noting, however, that the superiority of the bifactor model over the higher-order should be the object of a scientific debate that was not within the goals of the present study. In fact, previous evidence has shown that fit indices may be biased toward the bifactor model even when it is not the true model (e.g., [Murray & Johnson, 2013](#)). On a similar note, it could be argued that one of the advantages of a higher-order model, over the bifactor model, is that the g-factor does not

need to be partialled out (see [Reynolds & Keith, 2017](#), on this argument). In addition, we observed that when a Schmid-Leiman decomposed solution of the higher-order solution was adopted, the results were quite similar to those reported below for the bifactor model (see Supplemental material, Table S2). In favor of the bifactor model, however, it has been shown that the differences in fit between the bifactor vs. the higher-order models could be influenced by the violation of the proportionality constraint adopted in the higher-order model ([Gignac, 2016](#); but see also [Molenaar, 2016](#)). In this respect, when we calculated the coefficient of variation (CoV) for the g/s coefficient ratios using the procedure suggested by [Gignac \(2016\)](#), we observed a proportionality constraint violation ranging from moderate to large for the typically-developing group, and even stronger in the SLD group (see Supplemental material, Table S3). One may argue that this violation could be caused by the equality constraints that were imposed on the subtest loadings. However, in the model without any equality constraints, which was fitted using the American informed priors, violations were very large as well (Supplemental material, Table S3). The authors of the current report do not take any theoretical stance for the bifactor model over the higher-order one. However, for the specific purpose of our present paper, we believe that the use of the bifactor model could produce interesting and novel findings. These findings are also particularly relevant because, to the best of our knowledge, this is the first time that the proportionality constraint problem has been highlighted in a clinical population.

The posterior estimates of the model coefficients are reported, along with the percentage of overlapping between the Bayesian posterior distributions in the two groups, in [Fig. 1](#).

Figure 1 about here

The estimated common variance explained by the g-factor in the SLD group vs. the typically-developing group is reported in [Table 1](#) and also depicted in the Supplemental material (Fig. S1). As it can be seen, more than half of the common variance is left to be explained by the domain-specific factors in the SLD group, whereas the opposite is true for the typically-developing group. More precisely, there was very strong evidence that the g-factor explained <50% of common

variance in the SLD group ( $ER > 1000$ ), but  $>50\%$  of the common variance in the typically-developing group ( $ER > 1000$ ). There was also strong evidence that the g-factor explained a smaller portion of variance in the SLD than it did in the typically-developing group (overlapping  $<0.1\%$ ,  $ER > 1000.00$ ). Furthermore, the specific PR factor (with the g-factor partialled out) accounted for virtually no variance in each group. On the contrary, VC and PS explained good portions of common variance, especially in the SLD group.<sup>3</sup> Finally, [Table 1](#) shows that the model explained less than half of the total variance in both groups, namely 43.9% in the SLD group and 45.0% in the typically-developing group. Regarding the reliability of the global composite score, this was higher in the typically-developing group,  $\omega_h = 0.73$ , compared to the group with SLD  $\omega_h = 0.60$ . As for the PR, the reliabilities of this specific index approached 0 in the two groups. As for the other indices, the reliabilities ranged between 0.23 and 0.48 in the typically-developing group, and considerably higher in the SLD group, between 0.41 and 0.55. The posterior distributions of the loadings are represented in the Supplemental material, Fig. S2.

Table 1 about here

## 2.2. Additional analyses using American priors

The Bayesian priors were formalized as explained in the data analysis section (see also the Supplemental online material, Table S4). As the priors were calculated for 1617 observations, they were at least as strong as our data, thus having a potentially high leverage on the estimated values. Notably, the actual sample sizes were 1537 for the American SLD sample ([Styck & Watkins, 2016](#)), and 2200 for the Italian standardization sample (typically-developing group), but  $N = 1617$  was used for both groups for consistency with our SLD sample size.

Most estimated values were very similar, as before (for the sake of brevity, the full report is not presented here, but it can be found in the Supplemental online material, Table S5). Again, the majority of common variance was explained by the g-factor in the typically-developing group (67.0%) but not in the SLD group (42.1%), and there was strong evidence that the two figures are

different between-group, overlapping <0.1%, ER > 1000.00. The common variance explained by PR was greater compared to the study's previous results, but still small in both groups, i.e., 7.4% in SLD and 3.3% in typical development. The unique reliability of the indices remained nearly unchanged.

### 3. Discussion

In this study, we aimed to understand if the g-factor alone could sufficiently account for all the variance in the WISC-IV. Other specific indices (i.e., PR, VC, WM, and PS), were also evaluated in order to estimate the portion of the variance specifically accounted for by each of them in both children with SLD and with typical development. Results, obtained by analysing the scores collected from large samples of Italian children with SLD or without SLD at the WISC-IV subtests, confirmed that the structure of intelligence might present differences in the case of children with SLD compared to typically-developing children ([Giofrè & Cornoldi, 2015](#)). Furthermore, results showed that in children with typical development, the most prominent factor was clearly the g-factor. In fact, excluding the variance accounted for by error, the g-factor accounted for 60% of the common variance in these children. This result is not surprising and supports the prevalence of the g-factor compared to other factors in typically-developing children ([Gignac & Watkins, 2013](#); [Watkins, 2006](#)). Nevertheless, the specific factors were necessary, and the one-factor model, including the g-factor only, presented with a poor fit. Specific factors (i.e., PR, VC, WM, and PS), considered together, accounted for approximately 40% of the variance, suggesting that interpreting only the g-factor might result in a considerable loss of information even in the typically-developing population. It should be noted that we excluded cases with FSIQ <70 from our analysis, and this may be the reason that a smaller portion of variance was explained by the g-factor as compared with previous reports from the American population ([Watkins, 2006](#)), as dedifferentiation may be greater in low intelligence children ([Bremner, McTaggart, Saklofske, & Janzen, 2011](#); but see [Toffalini, Buono, Zagaria, Calcagni, and Cornoldi, 2019](#), for a different qualification). Even so, a



remarkable difference emerged between typically-developing children and those with SLD. In the latter group, the g-factor only accounted for 44% of the common variance, while the other main factors accounted for more than half of the common variance. Furthermore, these proportions changed very little when the American priors were used, as they were in line with the Italian results, confirming that our results were substantial and generalizable, despite the fact that SLD is related to difficulties in language processing and Italian and English present great linguistic differences. Overall, these findings seem to indicate that the main indices play a substantial role as indicators of specific domains above and beyond the g-factor, and they should be considered when assessing children with SLD.

The fact that the g-factor is relatively weak in children with SLD may be directly linked to the observation that SLD impairments occur in some specific domains, but not in others. For example, PR seems to be relatively intact in children with SLD, while other factors, for example WM, tend to be more or less severely impaired, and the implications of this observation are discussed further below. The present study showed that despite the fact that specific factors together explained the majority of common variance in the SLD population, not all of them had the same importance. The verbal domain and speed (i.e., VC and PS respectively) accounted for a large portion of variance of their subtests: VC accounted for 23.8% of common variance, which is notable if one considers that it could not explain >30%. PS accounted for 18.2% of common variance on a maximum of 20%. Moreover, WM also explained a respectable 13.1% out of a maximum of 20%. In contrast, the visuospatial domain (PR) accounted for nearly no variance, suggesting that the common variance of its subtests is fully captured by the g-factor, as suggested by previous studies (see [Keith, Fine, Taub, Reynolds, & Kranzler, 2006](#)). This finding contrasts with previous reports indicating that the VC index is relatively spared by SLD (e.g., [De Clercq-Quaegebeur et al., 2010](#); [Toffalini et al., 2017a](#)) and might have important implications.

The intellectual profiles of children with SLD seem to reflect the degree to which each area of functioning has been specifically spared or compromised by underlying neurodevelopmental

disorders, rather than a series of broad abilities that present specificities but also strongly reflect a single *g*-factor. It is well documented that children with various disabilities, including dyslexia, struggle, for example, with phonological working memory tasks (as in the case of the working memory tests included in the WISC-IV, e.g., [Swanson, 1999](#)), which are often used for the diagnosis of this disorder (see [Elliott & Grigorenko, 2014](#) for a review and [Melby-Lervåg, Lyster, & Hulme, 2012](#) for a meta-analysis). Furthermore, children with dyscalculia and dyslexia tend to struggle when the materials include numbers or letters; these stimuli are used in subtests included in the WISC-IV (e.g., the digit span test, the letter number sequencing, and the coding subtests), creating difficulties for children with SLD. Other tasks which supposedly measure processing speed, such as coding, tend to use visual abstract symbols, and it is well documented that children with dyscalculia, dysgraphia and even dyslexia tend to struggle with these tasks as well (e.g., [Gubbay & de Klerk, 1995](#); [Piazza et al., 2010](#); [Valdois, Bosse, & Tainturier, 2004](#)). As for symbol search, which is also included in the PSI and requires a rapid and careful analysis of a series of visual elements, the presence of subtle visual search deficits has often been associated with reading problems (e.g., [Casco, Tressoldi, & Dellantonio, 1998](#)). Finally, concerning the subtests (Similarities, Vocabulary and Comprehension) included in the VCI, the linguistic difficulties of many children with SLD could produce specific patterns of performance that are different from those observed in children with typical development and reflect to a lesser extent a general intellectual factor. For example, vocabulary tends to be impaired in both children with dyslexia and with reading comprehension problems, with a slow vocabulary development that is particularly evident in children with dyslexia (e.g., [Lyytinen, Eklund, & Lyytinen, 2005](#); [van Viersen et al., 2018](#)). The presence of these subtests in the WISC-IV could have profound implications both for the diagnosis of children with SLD and the comprehension of their relationship with *g*-factor.

It should be noted that some intelligence tests were initially developed to ascertain school difficulties. In particular, the [Binet and Simon \(1904\)](#) scale was commissioned by the French government to discriminate between children with or without school difficulties. In this respect, the

selection of subtests made by the authors was based on tasks that could discriminate well between children with or without school difficulties, independently from the loading on the g-factor, and the construction of the WISC scale was influenced by the Binet and Simon scale, but with the inclusion of the so-called performance scale subtests (see [Richardson, 2011](#) for an historical perspective). In fact, one main objective of intelligence test developers has not been to measure intelligence per se, but to provide tasks with greater discriminatory power to detect children with school problems. However, this may produce a series of problems and biases, as it happens, for example, for the inclusion of subtests such as Coding, with very low loadings on the g-factor.

In the present study we also found some interesting results concerning the omega estimates: specific factors tended to have higher omega scores in the SLD sample compared to the typically-developing group. On this respect, it is worth noting that the omega coefficient should not be interpreted using the conventional cut-offs, as these estimates tend to be considerably lower compared to other measures of reliability ([Gignac & Kretzschmar, 2017](#)). General criteria have been proposed for evaluating omega scores in terms of magnitude: relatively small  $<0.20$ ; moderate  $0.20$  to  $0.30$ ; and relatively large  $>0.30$  ([Gignac & Kretzschmar, 2017](#)). When these criteria are adopted, effects seem to be large in almost all cases in the SLD group, with the exception of PR having an extremely small score. In the group with typical development, the pattern was somehow different with large indices in the PS and VC, moderate in the WM case, and relatively small in the PR. Specifically, our findings showed that the PR omega scores were quite small in both groups, but particularly so and almost nil in children with SLD. In our view, this result should not be interpreted as confirmation that this index is not reliable, but as a demonstration that this particular index is rather a measure of the g-factor and very little, or nothing at all, remains as a reliable residual (specific) factor after partializing out the g-factor. This has important consequences, for example, for the evaluation of the g-factor in children with SLD. In this respect, we found that the VCI subtests were mainly related to their specific factor rather than to g. This finding suggests that indices such as the GAI, which includes both the PR and VC, might not be a good indicator of the

g-factor in children with SLD, because the variance included in the verbal comprehension test seems to reflect mainly specific influences and only weakly the g-factor. For this reason, it seems important to consider PRI and VCI separately.

With this paper, we also aimed to understand whether our results hold true for children from very different cultural and linguistic backgrounds (i.e., USA). To analyse this matter, we used a Bayesian approach. Italian and English are very different languages, as written Italian is highly transparent whereas English has a very opaque orthography. Therefore, children with SLD in Italy vs. Anglophone countries may face challenges for reasons that are partly or entirely different ([Ziegler & Goswami, 2005](#)). For example, the main focus of the assessment of children with dyslexia is speed for Italian speaking children while for English speaking children, the focal point is accuracy in decoding (e.g., [Oakhill, Cain, & Bryant, 2003](#); [Ziegler & Goswami, 2005](#)). In spite of this, when the available evidence from the American data was entered as a set of Bayesian priors, the estimates for these two language environments were remarkably similar.

Previous research on the American standardization sample reported that the g-factor explained the majority of common variance, i.e., 71.3%, thus leaving a very limited portion to be explained by domain-specific factors ([Watkins, 2006](#)). It should be noted that this figure was slightly higher than the 60.5% we found in the Italian typically-developing population; however, unlike we did in the present study, [Watkins \(2006\)](#) did not exclude the individuals with FSIQ <70 from his analysis. In contrast, with regard to American children with SLD, [Styck and Watkins \(2016\)](#), on the basis of data concerning a large sample of American children with SLD, reported that the g-factor explained 48% of common variance, thus leaving more than half to be explained by the domain-specific factors. Interestingly, they also found that VC and PS were relatively clearly defined as specific factors (in terms of the variance explained and reliability), whereas PR and WM were poor. In this same 2016 study, the authors mentioned as one of their limitations the fact that their sample, albeit large, was recruited in a single geographic location of the U.S. Therefore, the fact that our results appeared remarkably similar to those previously reported on the American

population, including loadings on the g-factor and on factors, omegas and all of the other parameters, demonstrates that certain differences between SLD and the general population tend to be quite robust across cultures, in spite of differences in context, language and the academic curriculum.

Although this was not the goal of the present study, we noted that our findings suggested a better fit of the bifactor model over the traditional higher-order model. In fact, this alleged superiority has been questioned, as some argue that the bifactor model may be overparametrized and attain a better fit even when it is not the true model that generated the data (e.g., [Murray & Johnson, 2013](#), see also [Bonifay & Cai, 2017](#); [Maydeu-Olivares & Coffman, 2006](#); [Murray & Johnson, 2013](#); and [Reise, Kim, Mansolf, & Widaman, 2016](#) for a similar argument). [Mansolf and Reise \(2017\)](#) showed that, when a higher-order model is true but mis-specified, i.e., there are unmodelled complexities such as correlated residuals, fit indices could erroneously favor the bifactor model. For these reasons, we must be cautious in selecting one over the other model. The use of the bifactor model in this paper was justified by the fact that the model provides valuable applications to understand the psychometrics of a scale and how its specific subscales represent distinct constructs (see [Bonifay, Lane, & Reise, 2016](#)), a question that we addressed across different samples belonging to very different populations. One important difference between the bifactor and the higher-order model, however, refers to the presence of the so-called “proportionality constraints” in the latter ([Gignac, 2016](#); but see also [Mansolf & Reise, 2017](#), for a more general framing of the issue). As explained in the Introduction, this means that in the higher-order model the relationship between each subtest and the superordinate g-factor is necessarily mediated by (i.e., constrained to be proportional to) the relationship between the first-order factors and the superordinate g-factor. Conversely, the bifactor model allowed us to estimate the relationship between each subtest and the g-factor without having this constraint. In fact, following [Gignac \(2016\)](#), we observed a violation of the proportionality constraint in both the TD and (even more so)

in the SLD population, irrespective of the use of equality constraints on the subtest loadings (Supplemental material, Table S3).

Despite the presence of novel and important findings, the present study also leaves unresolved issues, which should be considered in future research. The larger portion of the variance in the WISC-IV, which is one of the most prominent tests for the evaluation of intelligence, is not accounted for by the g-factor or by principal indices, but it is subtest-specific (i.e., error variance). This problem is reflected by the limited reliability of the specific indices in the typically-developing population, and of both the general and the specific composite scores in SLD. This result confirms that intelligence tests are imperfect indicators of broader intelligence behaviour ([Hunt, 2011](#)). Consistently with previous reports, the PR index accounts for a negligible percentage of the variance once the g-factor is controlled ([Toffalini et al., 2017b](#)). In addition, it should be noted that a new battery, the WISC-V, not available in many countries (e.g., Italy), was recently introduced in the U.S. This new battery comes with significant improvements, including for example a robust Nonverbal (Visual Spatial) Index and a g-fluid (Fluid Reasoning) Index; it would therefore be interesting to compare children with SLD to the normally developing control group using this battery. Finally, one should consider that only two cultures were examined. Despite the significance of the fact that the results were comparable in Italian and American children with SLD, it would also be interesting to evaluate if such findings are generalizable to various other cultures and contexts, particularly beyond Western contexts. The authors decided not to test exploratory solutions, compared to confirmatory ones. However, other rotation methods, e.g., BI-GEOMIN, could yield a factor solution from an unrestricted factor analysis that is highly similar to the unrestricted Schmid-Leiman (bi-factor analysis) ([Mansolf & Reise, 2016](#)), and future research could benefit from the use of this statistical approach.

Some implications for clinical practice with children with SLD can be also drawn from this paper. Although it is still the common practice in many countries including Italy, it is not advisable to use only the g-factor and ignore the other indices, because the g-factor considered in isolation

only accounts for a limited (if admittedly large) portion of the variance. At least VCI and PSI should be considered as distinct domains in children with SLD, whereas WMI could be of direct utility for supporting a diagnosis (e.g., [Giofrè et al., 2017](#)), and PRI could be considered alone as an indicator of *g*. In particular, it appears that VCI is very highly loaded on the *g*-factor, but also presents with very high portions of specific variance. This is particularly relevant also because children with SLD seem to be relatively unimpaired, or only to a minimal extent, in this factor (e.g., [De Clercq-Quaegebeur et al., 2010](#); [Toffalini et al., 2017a](#)). This finding is in accordance with previous evidence indicating that the verbal factor accounts for a significant portion of the variance in academic achievement after controlling for general intelligence ([Glutting, Watkins, Konold, & Mcdermott, 2006](#)). Furthermore, this finding seems to indicate that the PSI tend to be almost entirely independent from the *g*-factor. This is especially interesting because processing speed is frequently a weakness in the SLD population and therefore may have diagnostic significance along with WMI (e.g., [Giofrè et al., 2017](#)). Finally, practitioners working with children with typical and atypical development should be aware that intelligence, as measured by an intelligence battery, accounts only for a small part of intelligent behaviour. Focusing only on the *g*-factor, despite this factor being very powerful and reliable, is not advisable, particularly when working with children with SLD. At the same time, ignoring the *g*-factor in child assessment does not seem to be appropriate, as this factor appears to a very strong indicator, especially for children with typical development.

To conclude, we can confidently say that there are pros and cons to every aspect of the arguments made by lumpers and splitters. In considering typical development, it seems that the *g*-factor tends to prevail over the other factors, even if the variance accounted for by other principal indices is far from trivial and can hardly be neglected. Still, when considering children with SLD, the *g*-factor is very important, but the variance accounted for by all the other factors considered together is larger and should not be disregarded. Overall, these results seem to indicate that neither the lumpers nor the splitters are completely correct, and that rather, both the *g*-factor and principal

indices should be considered. It appears that this holds true for children with typical development and SLD across different cultures, as exemplified by the cases of Italy and of the U.S.



**Footnotes**

<sup>1</sup> With regards to the SLD subtypes according to the ICD-10 classification system, 350 children were diagnosed with reading disorder (F81.0); 202 children with spelling disorder (F81.1); 122 children with specific disorder of arithmetic skills (F81.2); 587 children with mixed disorder of scholastic skills (F81.3); and 356 children had SLD, unspecified, or other residual categories (F81.8, F81.9), or had received more than one distinct diagnosis within the F81 category. Any case presenting comorbid neuropsychological disorders other than SLD (e.g., attention-deficit hyperactivity disorder, developmental coordination disorder) was excluded from the sample through a preliminary screening.

<sup>2</sup> This was done because intellectual disability represents an exclusion criterion for SLD. At the same time, FSIQ in the borderline range, i.e., between 70 and 84, were not an exclusion criterion, and this condition represented 13.8% of the SLD sample (i.e., nearly the same as predicted by the normal distribution after cases below -2 SDs from the mean are removed, 13.9%). The correlation matrix calculated on the simulated subtest scores of the typically-developing group remained extremely similar to that of the Italian standardization sample (Orsini et al., 2012), although all correlations were between .01 and .04 points lower than those in the standardization sample (average = -.03), as a consequence of the removal of the lowest part of the FSIQ distribution.

<sup>3</sup> It should be noted that the common variance explained by the g-factor could theoretically reach 100%, whereas common variance explained by VC or PR cannot exceed 30% (only three out of ten subtests involved), and common variance explained by WM or PS cannot exceed 20% (only two out of ten subtests involved).

### References

- Beaujean, A. A. (2014). *Latent variable modelling using R: a step by step guide*. Routledge, Taylor & Francis Group, New York.
- Binet, A., & Simon, T. (1904). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique*, *11*, 191–244.
- Bonifay, W., & Cai, L. (2017). Is the bifactor model a better model or is it just better at modeling implausible responses? Application of iteratively reweighted least squares to the Rosenberg self-esteem scale. *Multivariate Behavioral Research*, *52*(4), 465–484.  
doi:10.1080/00273171.2017.1309262
- Bonifay, W., Lane, S. P., & Reise, S. P. (2016). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, *5*, 184-186.  
doi:10.1177/2167702616657069
- Bremner, D., McTaggart, B., Saklofske, D. H., & Janzen, T. (2011). WISC-IV GAI and CPI in psychoeducational assessment. *Canadian Journal of School Psychology*, *26*, 209-219.  
doi:10.1177/0829573511419090
- Canivez, G. L. (2014). Construct validity of the WISC-IV with a referred sample: Direct versus indirect hierarchical structures. *School Psychology Quarterly*, *29*(1), 38–51.  
doi:10.1037/spq0000032
- Canivez, G. L., Watkins, M. W., Good, R., James, K., & James, T. (2017). Construct validity of the Wechsler intelligence scale for children – Fourth UK Edition with a referred Irish sample: Wechsler and Cattell–Horn–Carroll model comparisons with 15 subtests. *British Journal of Educational Psychology*, *87*(3), 383–407. doi:10.1111/bjep.12155
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.

- Casco, C., Tressoldi, P. E., & Dellantonio, A. (1998). Visual selective attention and reading efficiency are related in children. *Cortex*, *34*(4), 531–546. [https://doi.org/10.1016/S0010-9452\(08\)70512-4](https://doi.org/10.1016/S0010-9452(08)70512-4)
- Chen, H., Keith, T. Z., Weiss, L., Zhu, J., & Li, Y. (2010). Testing for multigroup invariance of second-order WISC-IV structure across China, Hong Kong, Macau, and Taiwan. *Personality and Individual Differences*, *49*(7), 677–682. doi:10.1016/j.paid.2010.06.004
- Chen, H., & Zhu, J. (2012). Measurement invariance of WISC-IV across normative and clinical samples. *Personality and Individual Differences*, *52*(2), 161–166. doi:10.1016/j.paid.2011.10.006
- Cornoldi, C., Giofrè, D., Orsini, A., & Pezzuti, L. (2014). Differences in the intellectual profile of children with intellectual vs. learning disability. *Research in Developmental Disabilities*, *35*(9), 2224–2230. doi:10.1016/j.ridd.2014.05.013
- D'Angiulli, A., & Siegel, L. S. (2003). Cognitive functioning as measured by the WISC-R: Do children with learning disabilities have distinctive patterns of performance? *Journal of Learning Disabilities*, *36*, 48–58. doi:10.1177/00222194030360010601
- De Clercq-Quaegebeur, M., Casalis, S., Lemaitre, M.-P., Bourgois, B., Getto, M., & Vallée, L. (2010). Neuropsychological profile on the WISC-IV of French children with dyslexia. *Journal of Learning Disabilities*, *43*(6), 563–574. doi:10.1177/0022219410375000
- Elliott, J., & Grigorenko, E. (2014). Explanations at the cognitive level. In *The Dyslexia Debate* (Cambridge Studies in Cognitive and Perceptual Development, pp. 42-87). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139017824.004
- Fiorello, C. A., Hale, J. B., Holdnack, J. A., Kavanagh, J. A., Terrell, J., & Long, L. (2007). Interpreting intelligence test results for children with disabilities: Is global intelligence relevant? *Applied Neuropsychology*, *14*(1), 2–12. doi:10.1080/09084280701280338

- Fiorello, C. A., Hale, J. B., Mcgrath, M., Ryan, K., & Quinn, S. (2001). IQ interpretation for children with flat and variable test profiles. *Learning and Individual Differences, 13*(2), 115–125. doi:10.1016/S1041-6080(02)00075-4
- Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV Assessment*. New York, NY: Wiley.
- Fletcher, J. M., Lyonm, G. R., Barnes, M., Stuebing, K. K., Francis, D. J., Olson, R. K., ... Shaywitz, B. A. (2002). *Classification of learning disabilities: An evidence-based evaluation*. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 185–250). Mahwah, NJ: Lawrence Erlbaum.
- Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence, 55*, 57–68. doi:10.1016/j.intell.2016.01.006
- Gignac, G. E., & Kretzschmar, A. (2017). Evaluating dimensional distinctness with correlated-factor models: Limitations and suggestions. *Intelligence, 62*, 138–147. doi:10.1016/j.intell.2017.04.001
- 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. doi:10.1080/00273171.2013.804398
- Giofrè, D., & Cornoldi, C. (2015). The structure of intelligence in children with specific learning disabilities is different as compared to typically development children. *Intelligence, 52*, 36–43. doi:10.1016/j.intell.2015.07.002
- Giofrè, D., Toffalini, E., Altoè, G., & Cornoldi, C. (2017). Intelligence measures as diagnostic tools for children with specific learning disabilities. *Intelligence, 61*, 140–145. doi:10.1016/j.intell.2017.01.014
- Glutting, J. J., Watkins, M. W., Konold, T. R., & Mcdermott, P. A. (2006). Distinctions without a difference: Math achievement on the WIAT-II. *The Journal of Special Education, 40*, 103–114. doi:10.1177/00224669060400020101

- Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2013). Further insights on the French WISC-IV factor structure through Bayesian structural equation modelling. *Psychological Assessment, 25*(2), 496–508. doi:10.1037/a0030676
- Gomez, R., Vance, A., & Watson, S. D. (2016). Structure of the Wechsler intelligence scale for children – fourth edition in a group of children with ADHD. *Frontiers in Psychology, 7*, 1–11. doi:10.3389/fpsyg.2016.00737
- Gottfredson, L. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24*(1), 13–23. doi:10.1016/S0160-2896(97)90011-8
- Gubbay, S. S., & de Klerk, N. H. (1995). A study and review of developmental dysgraphia in relation to acquired dysgraphia. *Brain and Development, 17*(1), 1–8. doi:10.1016/0387-7604(94)00110-J
- Guilford, J. P. (1967). *The nature of human intelligence*. New York, NY: McGraw-Hill.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*(5), 253–270. doi:10.1037/h0023816
- Keith, T. Z. (2005). *Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests*. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 581–614). New York, NY: Guilford Press.
- Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler intelligence scale for children–fourth edition: what does it measure? *School Psychology Review, 35*, 108–127.
- Kruschke, J. K., & Liddell, T. M. (2015). The Bayesian new statistics: Two historical trends converge. *SSRN Electronic Journal, 2606016*.

- Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence, 33*(4), 393–416.  
doi:10.1016/j.intell.2004.12.002
- Lecerf, T., Rossier, J., Favez, N., Reverte, I., & Coleaux, L. (2010). The four- vs. alternative six-factor structure of the French WISC-IV. *Swiss Journal of Psychology, 69*(4), 221–232.  
doi:10.1024/1421-0185/a000026
- Lyytinen, P., Eklund, K., & Lyytinen, H. (2005). Language development and literacy skills in late-talking toddlers with and without familial risk for dyslexia. *Annals of Dyslexia, 55*(2), 166–192. doi:10.1007/s11881-005-0010-y
- Mansolf, M., & Reise, S. P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence, 61*, 120–129. doi:10.1016/j.intell.2017.01.012
- Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods, 11*(4), 344–362. doi:10.1037/1082-989X.11.4.344
- Mayes, S. D., & Calhoun, S. L. (2007). Wechsler Intelligence Scale for Children-Third and -Fourth Edition predictors of academic achievement in children with attention-deficit/hyperactivity disorder. *School Psychology Quarterly, 22*(2), 234–249. doi:10.1037/1045-3830.22.2.234
- Mayes, S. D., & Calhoun, S. L. (2008). WISC-IV and WIAT-II profiles in children with high-functioning autism. *Journal of Autism and Developmental Disorders, 38*(3), 428–439.  
doi:10.1007/s10803-007-0410-4
- McGill, R. J., & Canivez, G. L. (2017). Confirmatory factor analyses of the WISC-IV Spanish core and supplemental subtests: Validation evidence of the Wechsler and CHC models. *International Journal of School & Educational Psychology, 1–13*.  
doi:10.1080/21683603.2017.1327831
- Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138*(2), 322–352.  
doi:10.1037/a0026744

- Merkle, E. C. & Rosseel, Y. (2018). blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, 85. doi:10.18637/jss.v085.i04
- Molenaar, D. (2016). On the distortion of model fit in comparing the bifactor model and the higher-order factor model. *Intelligence*, 57, 60–63. doi:10.1016/j.intell.2016.03.007
- Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus the higher-order models of human cognitive ability structure. *Intelligence*, 41, 407-422. doi:10.1016/j.intell.2013.06.004
- Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes*, 18, 443-468. doi: 10.1080/01690960344000008
- Oliveras-Rentas, R. E., Kenworthy, L., Roberson, R. B., Martin, A., & Wallace, G. L. (2012). WISC-IV profile in high-functioning autism spectrum disorders: Impaired processing speed is associated with increased autism communication symptoms and decreased adaptive communication abilities. *Journal of Autism and Developmental Disorders*, 42(5), 655–664. doi:10.1007/s10803-011-1289-7
- Pastore, M. (2018). Overlapping: a R package for estimating overlapping in empirical distributions. *The Journal of Open Source Software*, 3:1023.
- Peng, P., & Fuchs, D. (2016). A Meta-Analysis of Working Memory Deficits in Children With Learning Difficulties: Is There a Difference Between Verbal Domain and Numerical Domain?. *Journal of Learning Disabilities*, 49, 3-20. doi:10.1177/0022219414521667
- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., ... Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, 116(1), 33–41. doi:10.1016/j.cognition.2010.03.012
- Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3<sup>rd</sup> International*

*Workshop on Distributed Statistical Computing (DSC 2003)*. Technische Universität Wien, Vienna, Austria.

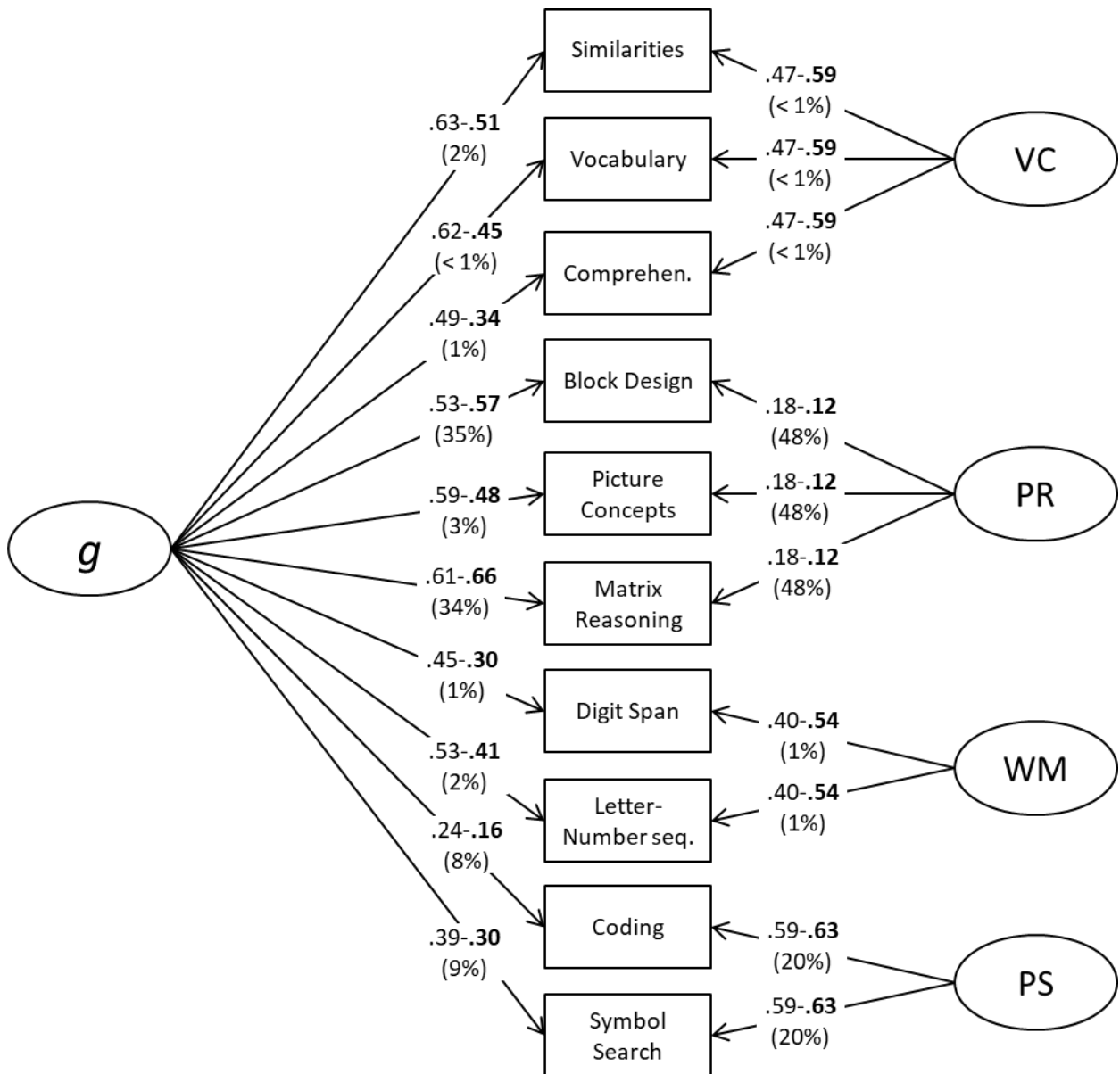
- Poletti, M. (2016). WISC-IV intellectual profiles in Italian children with specific learning disorder and related impairments in reading, written expression, and mathematics. *Journal of Learning Disabilities, 49*(3), 320–335. doi:10.1177/0022219414555416
- Provazza, S., Giofrè, D., Adams, A-M., & Roberts, D. J. (2019). *Word length effects in adult participants with developmental dyslexia and typical readers*. Manuscript submitted for publication.
- R Core Team (2018). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111-163.
- Reynolds, M. R., & Keith, T. Z. (2017). Multi-group and hierarchical confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition: What does it measure? *Intelligence, 62*, 31-47. doi:10.1016/j.intell.2017.02.005
- Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*, 129–140. doi:10.1080/00223891.2012.725437
- Reise, S. P., Kim, D. S., Mansolf, M., & Widaman, K. F. (2016). Is the bifactor model a better model or is it just better at modeling implausible responses? Application of iteratively reweighted least squares to the Rosenberg self-esteem scale. *Multivariate Behavioral Research*. doi:10.1080/00273171.2016.1243461
- Reverte, I., Golay, P., Favez, N., Rossier, J., & Lecerf, T. (2015). Testing for multigroup invariance of the WISC-IV structure across France and Switzerland: Standard and CHC models. *Learning and Individual Differences, 40*, 127–133. doi:10.1016/j.lindif.2015.03.015



- Richardson, J. T. E. (2011). *Howard Andrew Knox: Pioneer of intelligence testing at Ellis Island*. New York, NY: Columbia University Press.
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22, 53-61. doi:10.1007/BF02289209
- Styck, K. M., & Watkins, M. W. (2016). Structural validity of the WISC-IV for students with learning disabilities. *Journal of Learning Disabilities*, 49(2), 216–224. doi:10.1177/0022219414539565
- Sumner, E., Pratt, M. L., & Hill, E. L. (2016). Examining the cognitive profile of children with Developmental Coordination Disorder. *Research in Developmental Disabilities*, 56, 10–17. doi:10.1016/j.ridd.2016.05.012
- Swanson, L. L. (1999). Reading comprehension and working memory in learning-disabled readers: Is the phonological loop more important than the executive system? *Journal of Experimental Child Psychology*, 72(1), 1–31. doi:10.1006/jecp.1998.2477
- Thaler, N. S., Bello, D. T., & Etcoff, L. M. (2013). WISC-IV Profiles are associated with differences in symptomatology and outcome in children with ADHD. *Journal of Attention Disorders*, 17(4), 291–301. doi:10.1177/1087054711428806
- Thorndike, E. L. (1903). *Educational psychology*. New York, NY: Lemcke & Buechner.
- Toffalini, E., Buono, S., Zagaria, T., Calcagnì, A., & Cornoldi, C. (2019). Using Z and age-equivalent scores to address WISC-IV floor effects for children with intellectual disability. *Journal of Intellectual Disability Research*. doi:10.1111/jir.12589
- Toffalini, E., Giofrè, D., & Cornoldi, C. (2017a). Strengths and weaknesses in the intellectual profile of different subtypes of specific learning disorder. *Clinical Psychological Science*, 5(2), 402–409. doi:10.1177/2167702616672038
- Toffalini, E., Giofrè, D., & Cornoldi, C. (2017b). Pros and cons of using intelligence batteries for the study of clinical populations: A response to Beaujean (2017). *Clinical Psychological Science*, 5(5), 878–879. doi:10.1177/2167702617720716

- Valdois, S., Bosse, M.-L., & Tainturier, M.-J. (2004). The cognitive deficits responsible for developmental dyslexia: Review of evidence for a selective visual attentional disorder. *Dyslexia, 10*(4), 339–363. doi:10.1002/dys.284
- van de Shoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: applications to developmental research. *Child Development, 85*, 842-860. doi:10.1111/cdev.12169
- van Viersen, S., de Bree, E. H., Zee, M., Maassen, B., van der Leij, A., & de Jong, P. F. (2018). Pathways Into Literacy: The Role of Early Oral Language Abilities and Family Risk for Dyslexia. *Psychological Science, 29*(3), 418–428. doi:10.1177/0956797617736886
- Vandekerckhove, J., Rouder, J. N., & Kruschke, J. K. (Eds.). (2018). Bayesian Inference for Psychology [Special issue]. *Psychonomic Bulletin & Review, 25*, 1-481.
- Vernon, P. E. (1950). *The Structure of Human Abilities*. New York: Wiley.
- Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research, 14*, 867-897.
- Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children—Fourth edition. *Psychological Assessment, 18*, 123-125. doi:10.1037/1040-3590.18.1.123
- Watkins, M. W., Canivez, G. L., James, T., James, K., & Good, R. (2013). Construct validity of the WISC–IV UK with a large referred Irish sample. *International Journal of School & Educational Psychology, 1*(2), 102–111. doi:10.1080/21683603.2013.794439
- Watkins, M. W., Glutting, J. J., & Lei, P.-W. (2007). Validity of the full-scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology, 14*(1), 13–20. doi:10.1080/09084280701280353
- Watkins, M. W., & Smith, L. G. (2013). Long-term stability of the Wechsler Intelligence Scale for Children - fourth edition. *Psychological Assessment, 25*(2), 477–483. doi:10.1037/a0031653

- Wechsler, D. (2003). *Wechsler intelligence scale for children—fourth edition technical and interpretive manual*. San Antonio, TX: Psychological Corporation
- Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WISC-IV and clinical validation of the four- and five-factor interpretative approaches. *Journal of Psychoeducational Assessment*, 31(2), 114–131. doi:10.1177/0734282913478032
- Willis, J. O., Dumont, R., & Kaufman, A. S. (2011). *Factor-analytic models of intelligence*. R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 39–57). New York: Cambridge University Press.
- World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva, Switzerland.



**Figure 1.** Bifactor structure of the WISC-IV for 1,617 with a specific learning disorder (parameter on right, in bold) and 1,617 simulated typically-developing children (with  $IQ \geq 70$ ; parameter on left). Percentages of overlapping between the Bayesian posterior distributions in the two groups are reported in parentheses.

**Table 1.** Estimated values in the Wechsler Intelligence Scale for Children–Fourth Edition among 1,617 children with a specific learning disorder and an equal number of simulated typically-developing children (with  $IQ \geq 70$ ), using a set of non-informed priors.

<b>Group: Specific Learning Disorder (N = 1,617)</b>												
Subtest	General		VC		PR		WM		PS		h <sup>2</sup>	u <sup>2</sup>
	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var	<i>B</i>	Var	<i>b</i>	Var		
SI	0.508	0.258	0.590	0.348							0.606	0.394
VO	0.446	0.199	0.590	0.348							0.547	0.453
CO	0.336	0.113	0.590	0.348							0.461	0.539
BD	0.570	0.325			0.125	0.016					0.341	0.659
PCn	0.477	0.228			0.125	0.016					0.243	0.757
MR	0.654	0.428			0.125	0.016					0.443	0.557
DS	0.304	0.092					0.536	0.287			0.380	0.620
LN	0.407	0.166					0.536	0.287			0.453	0.547
CD	0.155	0.024							0.631	0.398	0.422	0.578
SS	0.304	0.092							0.631	0.398	0.491	0.509
% total variance		19.2		10.4		0.5		5.7		8.0	43.9	56.1
% common variance		43.9		23.9		1.0		13.1		18.1		
<i>wh</i>		0.598		0.507		0.028		0.407		0.551		
<b>Group: Typical Development (simulated from the standardization sample, with FSIQs <math>\geq 70</math>; N = 1,617)</b>												
Subtest	General		VC		PR		WM		PS		h <sup>2</sup>	u <sup>2</sup>
	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var	<i>B</i>	Var	<i>b</i>	Var		
SI	0.628	0.394	0.471	0.222							0.616	0.384
VO	0.623	0.388	0.471	0.222							0.610	0.390
CO	0.485	0.235	0.471	0.222							0.457	0.543
BD	0.531	0.282			0.178	0.032					0.314	0.686
PCn	0.591	0.349			0.178	0.032					0.381	0.619
MR	0.614	0.377			0.178	0.032					0.409	0.591
DS	0.454	0.206					0.399	0.159			0.365	0.635
LN	0.533	0.284					0.399	0.159			0.443	0.557

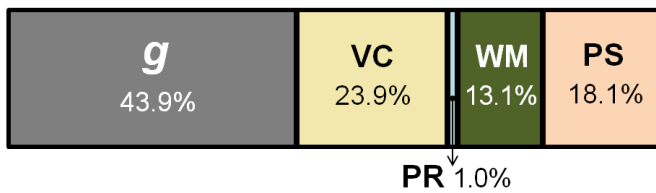
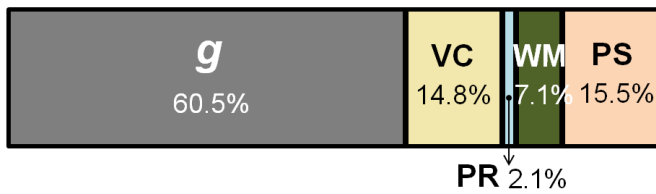
CD	0.242	0.059				0.590	0.348	0.407	0.593
SS	0.385	0.148				0.590	0.348	0.496	0.504
% total variance		27.2	6.7	1.0	3.2		7.0	45.0	55.0
% common variance		60.5	14.8	2.1	7.1		15.5		
$\omega h$		0.725	0.315	0.060	0.228		0.483		

*Note.*  $b$  = estimated standardized loading of subtest on factor; BD = Block Design; CD = Coding; CO = Comprehension; DS = Digit Span;  $h^2$  = communality; LN = Letter-Number sequencing; MR = Matrix Reasoning; PCn = Picture Concepts; PR = Perceptual Reasoning factor; PS = Processing Speed factor; SI = Similarities; SS = Symbol Search;  $u^2$  = uniqueness; Var = proportion of variance of subtest explained by the factor; VC = Verbal Comprehension factor; VO = Vocabulary; WM = Working Memory factor.

**Supplemental Material:**

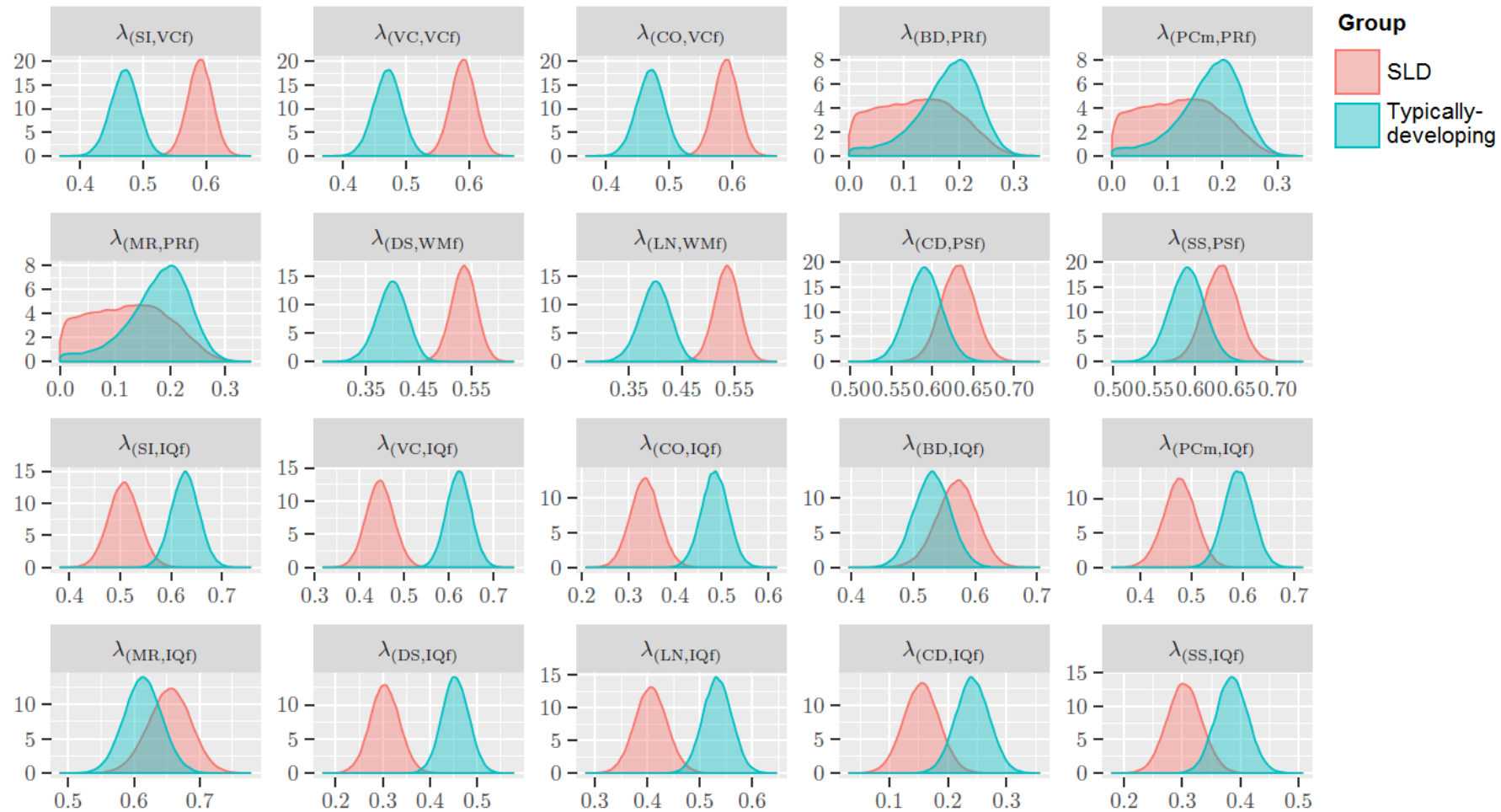
**Lumpers vs. splitters: Intelligence in children with specific learning disorders**

## Specific learning disorder

Typical development (with IQ  $\geq$  70)

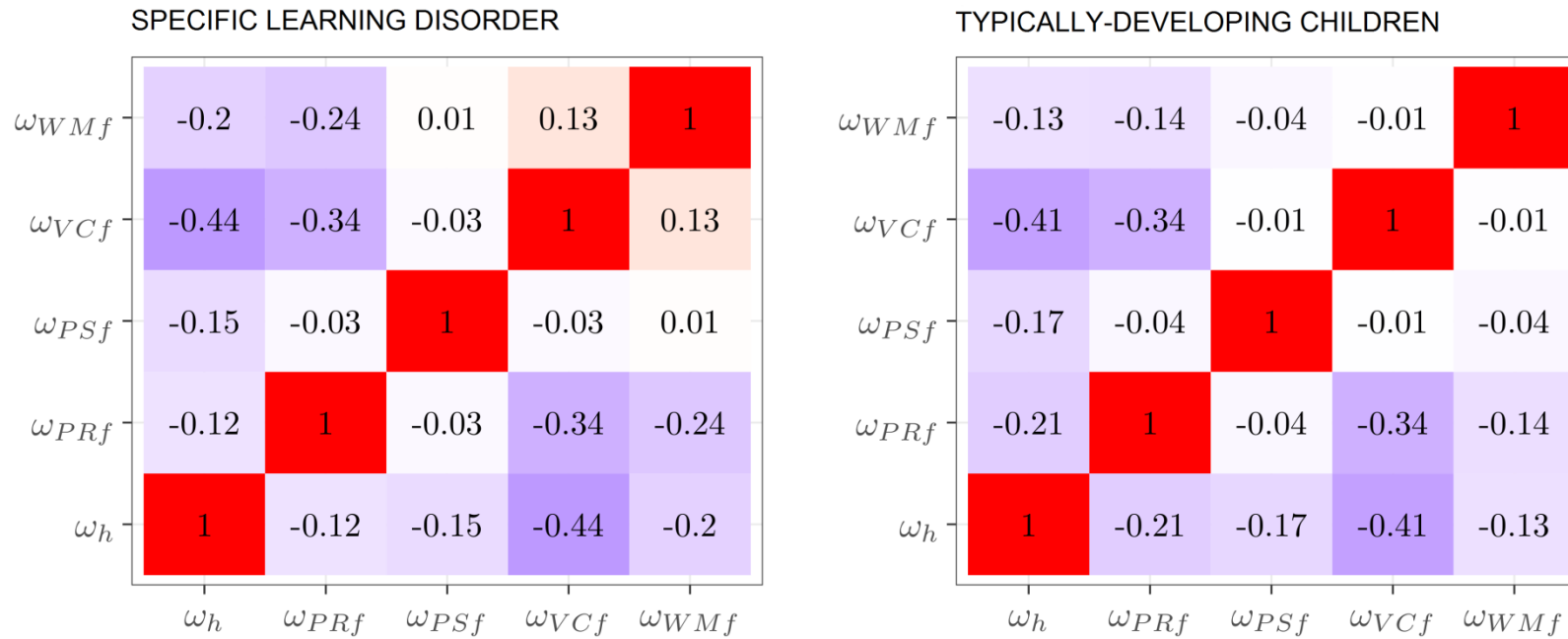
**Figure S1.** Percentage of common variance of the 10 WISC-IV subtests explained by the *g* factor and the specific factors in 1,617 children with a specific learning disorder and 1,617 simulated typically-developing children.





**Figure S2.** Posterior distributions of the standardized loading coefficients (using non-informative priors).

*Note.* BD = Block Design; CD = Coding; CO = Comprehension; DS = Digit Span; IQf = Intelligence Quotient (general factor); LN = Letter-Number sequencing; MR = Matrix Reasoning; PCn = Picture Concepts; PRf = Perceptual Reasoning factor; PSf = Processing Speed factor; SI = Similarities; SS = Symbol Search; VCf = Verbal Comprehension factor; VC = Vocabulary; WMf = Working Memory factor.



**Figure S3.** Bivariate correlations between the  $\omega$  measures of the g-factor and the specific factors calculated on the posterior distributions. As expected, the  $\omega$  of the g-factor is negatively correlated with the  $\omega$  of the specific factors (and of VCI in particular). This indicates that a very high internal consistency of the g-factor is incompatible with a strong distinctness of the specific factors. This holds true for both the specific learning disorder and the typically-developing population.

**Table S1.** Means (and standard deviations) of the WISC-IV subtests and indices in Italian children with SLD ( $N = 1,617$ ) and typically-developing children (after  $FSIQ < 70$  are removed). The values for typically-developing children are estimated through simulations on 1,000,000 cases.

<i>Variable</i>	SLD	Typically-developing (with $FSIQ \geq 70$ )
<i>Full-Scale IQ</i>	98.32 (12.44)	100.68 (14.41)
<i>Factorial indexes</i>		
Verbal Comprehension Index	103.15 (14.81)	100.68 (14.41)
Perceptual Reasoning Index	104.30 (14.24)	100.69 (14.40)
Working Memory Index	90.22 (13.02)	100.59 (14.56)
Processing Speed Index	93.14 (14.11)	100.49 (14.70)
<i>Subtests</i>		
Similarities	10.27 (2.90)	10.12 (2.91)
Vocabulary	10.38 (2.82)	10.12 (2.91)
Comprehension	10.93 (3.20)	10.11 (2.93)
Block design	10.36 (2.82)	10.10 (2.93)
Picture Concepts	11.02 (2.93)	10.11 (2.93)
Matrix reasoning	10.69 (3.03)	10.11 (2.93)
Digit span	8.18 (2.55)	10.10 (2.94)
Letter-number seq.	8.56 (2.56)	10.10 (2.93)
Coding	8.37 (2.87)	10.07 (2.97)
Symbol search	9.32 (2.75)	10.09 (2.95)

**Table S2.** Schmid-Leiman decomposed solution of the higher-order model. Estimated values in the Wechsler Intelligence Scale for Children–Fourth Edition for SLD vs typically-developing children using a set of non-informed priors.

<b>Group: Specific Learning Disorder (N = 1617)</b>												
Subtest	General		VC		PR		WM		PS		h <sup>2</sup>	u <sup>2</sup>
	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var		
SI	0.474	0.225	0.594	0.353							0.578	0.422
VO	0.485	0.235	0.607	0.368							0.604	0.396
CO	0.402	0.162	0.503	0.253							0.415	0.585
BD	0.552	0.305			0.250	0.063					0.367	0.633
PCn	0.456	0.208			0.206	0.042					0.250	0.750
MR	0.635	0.403			0.287	0.082					0.486	0.514
DS	0.322	0.104					0.447	0.200			0.303	0.697
LN	0.435	0.189					0.605	0.366			0.555	0.445
CD	0.174	0.030							0.477	0.228	0.258	0.742
SS	0.300	0.090							0.822	0.676	0.766	0.234
% total variance		19.5		9.7		1.9		5.7		9.0	45.8	54.2
% common variance		42.6		21.3		4.1		12.4		19.7		
$\omega_{h/s}$		0.606		0.471		0.107		0.392		0.584		
<b>Group: Typical Development (simulated from the standardization sample, without IQ &lt; 70; N = 1617)</b>												
Subtest	General		VC		PR		WM		PS		h <sup>2</sup>	u <sup>2</sup>
	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var		
SI	0.592	0.350	0.467	0.218							0.569	0.431
VO	0.636	0.404	0.502	0.252							0.657	0.344
CO	0.523	0.274	0.413	0.171							0.444	0.556
BD	0.541	0.293			0.189	0.036					0.328	0.672
PCn	0.598	0.358			0.209	0.044					0.401	0.599
MR	0.626	0.392			0.219	0.048					0.440	0.560
DS	0.455	0.207					0.363	0.132			0.339	0.661
LN	0.538	0.289					0.430	0.185			0.474	0.526
CD	0.242	0.059							0.467	0.218	0.277	0.723

SS	0.385	0.148				0.743	0.552	0.700	0.300
% total variance		27.7	6.4	1.3	3.2		7.7	46.3	53.7
% common variance		59.9	13.8	2.8	6.8		16.6		
$\omega_{h/s}$		0.730	0.303	0.071	0.224		0.508		

*Note.* *b* = estimated standardized loading of subtest on factor; BD = Block Design; CD = Coding; CO = Comprehension; DS = Digit Span;  $h^2$  = communality; LN = Letter-Number sequencing; MR = Matrix Reasoning; PCn = Picture Concepts; PR = Perceptual Reasoning factor; PS = Processing Speed factor; SI = Similarities; SS = Symbol Search;  $u^2$  = uniqueness; Var = proportion of variance of subtest explained by the factor; VC = Verbal Comprehension factor; VO = Vocabulary; WM = Working Memory factor.

**Table S3.** Violation of the proportionality constraint in the SLD and the typically-developing (TD) populations. Using the higher-order model, the proportionality constraint is imposed by definition, therefore the coefficient of variation (CoV; see Gignac, 2016) is constrained to zero. The same coefficient in the bifactor model (here fitted both using non-informative and American-sample-based informative priors; see below), quantifies the degree of proportionality constraint violation. Gignac (2016) interprets  $\text{CoV} = .37$  to  $.49$  as “moderate violation”. In the cases presented below, the violation seems always more than moderate, especially in the SLD population, and irrespective of the prior used.

Subtest	Using non-informative priors				Using the American priors (see below)	
	SLD population		TD population		SLD population	TD population
	Higher-Order (S-L)	Bifactor	Higher-Order (S-L)	Bifactor	Bifactor	Bifactor
	<i>g/s</i>	<i>g/s</i>	<i>g/s</i>	<i>g/s</i>	<i>g/s</i>	<i>g/s</i>
SI	0.80	0.86	1.27	1.33	1.05	1.87
VO	0.80	0.76	1.27	1.32	0.69	1.08
CO	0.80	0.57	1.27	1.03	0.68	1.03
BD	2.21	4.56	2.86	2.98	1.87	2.54
PCn	2.21	3.82	2.86	3.32	57.11	6.57
MR	2.21	5.23	2.86	3.45	1.39	1.43
DS	0.72	0.57	1.25	1.14	0.55	1.18
LN	0.72	0.76	1.25	1.34	0.82	1.37
CD	0.36	0.25	0.52	0.41	0.23	0.40
SS	0.36	0.48	0.52	0.65	0.48	0.66
CoV	0.00	1.02	0.00	0.65	2.62	1.58

*Note.* CoV = coefficient of variation (summed across the three first-order domains of Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed). *g/s* = *g* loading divided by the corresponding first-order specific factor loading. S-L = Schmid-Leiman decomposed solution. See Gignac (2016) for details.

*Subtests:* BD = Block Design; CD = Coding; CO = Comprehension; DS = Digit Span; LN = Letter-Number sequencing; MR = Matrix Reasoning; PCn = Picture Concepts; PR = Perceptual Reasoning factor; PS = Processing Speed factor; SI = Similarities; SS = Symbol Search; VO = Vocabulary;

**Table S4.** Estimated model coefficients defined on the available American data for the specific learning disorder (SLD) sample, and typically-developing sample.

Loading coefficient	SLD sample		Typically-developing sample	
	M	SD	M	SD
VC → SI	.530	.018	.464	.018
VC → VO	.740	.018	.464	.018
VC → CO	.500	.018	.464	.018
PR → BD	.270	.029	.250	.029
PR → PCn	.110	.029	.250	.029
PR → MR	.560	.029	.250	.029
WM → DS	.420	.025	.379	.025
WM → LN	.380	.025	.379	.025
PS → CD	.600	.021	.509	.021
PS → SS	.650	.021	.509	.021
g → SI	.500	.024	.713	.024
g → VO	.530	.023	.729	.023
g → CO	.450	.025	.614	.025
g → BD	.540	.025	.667	.025
g → PCn	.560	.026	.578	.026
g → MR	.610	.024	.696	.024
g → DS	.420	.025	.542	.025
g → LN	.510	.024	.639	.024
g → CD	.210	.026	.461	.026
g → SS	.460	.025	.588	.025

*Note.* BD = Block Design; CD = Coding; CO = Comprehension; DS = Digit Span; g = general factor; LN = Letter-Number sequencing; MR = Matrix Reasoning; PCn = Picture Concepts; PR = Perceptual Reasoning factor; PS = Processing Speed factor; SI = Similarities; SS = Symbol Search; VC = Verbal Comprehension factor; VO = Vocabulary; WM = Working Memory factor.

**Table S5.** Estimated values in the Wechsler Intelligence Scale for Children–Fourth Edition among 1617 children with a specific learning disorder and an equal number of simulated typically-developing children (with  $IQ \geq 70$ ), calculated using the informed priors defined on available American data.

<b>Group: Specific Learning Disorder (N = 1617)</b>												
Subtest	General		VC		PR		WM		PS		h <sup>2</sup>	u <sup>2</sup>
	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var		
SI	0.542	0.294	0.519	0.269							0.563	0.437
VC	0.491	0.241	0.698	0.487							0.728	0.272
CO	0.419	0.176	0.512	0.262							0.438	0.562
BD	0.524	0.275			0.260	0.068					0.342	0.658
PCn	0.523	0.274			0.092	0.008					0.282	0.718
MR	0.599	0.359			0.554	0.307					0.666	0.334
DS	0.389	0.151					0.600	0.360			0.511	0.489
LN	0.488	0.238					0.651	0.424			0.662	0.338
CD	0.202	0.041							0.654	0.428	0.469	0.531
SS	0.382	0.146							0.632	0.399	0.545	0.455
% total variance		21.9		10.2		3.8		7.8		8.3	52.1	47.9
% common variance		42.1		19.6		7.4		15.1		15.9		
$\omega_{h/s}$		0.637		0.469		0.157		0.495		0.555		
<b>Group: Typical Development (simulated from the standardization sample, without IQ &lt; 70; N = 1617)</b>												
Subtest	General		VC		PR		WM		PS		h <sup>2</sup>	u <sup>2</sup>
	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var	<i>b</i>	Var		



LUMPERS VS. SPLITTERS IN SLD AND TYPICAL DEVELOPMENT

49

SI	0.703	0.494	0.403	0.162				0.657	0.343
VC	0.694	0.482	0.517	0.267				0.749	0.251
CO	0.564	0.318	0.474	0.225				0.543	0.457
BD	0.605	0.366			0.232	0.054		0.420	0.580
PCn	0.602	0.362			0.222	0.049		0.412	0.588
MR	0.661	0.437			0.266	0.071		0.508	0.492
DS	0.516	0.266					0.384	0.147	0.414
LN	0.603	0.364					0.383	0.147	0.510
CD	0.371	0.138						0.546	0.298
SS	0.509	0.259						0.546	0.298
% total variance		34.9		6.5		1.7		2.9	6.0
% common variance		67.0		12.6		3.3		5.7	11.5
$\omega_{h/s}$		0.790		0.284		0.091		0.202	0.401

*Note.* *b* = estimated standardized loading of subtest on factor; BD = Block Design; CD = Coding; CO = Comprehension; DS = Digit Span;  $h^2$  = communality; LN = Letter-Number sequencing; MR = Matrix Reasoning; PCn = Picture Concepts; PR = Perceptual Reasoning factor; PS = Processing Speed factor; SI = Similarities; SS = Symbol Search;  $u^2$  = uniqueness; Var = proportion of variance of subtest explained by the factor; VC = Verbal Comprehension factor; VO = Vocabulary; WM = Working Memory factor.