

Enhancing the Potential of IDS-2: Proposal of Two New Clinically Sensitive Indices of Intellectual Ability

Enrico Toffalini, Cesare Cornoldi
Department of General Psychology, University of Padua



enricotoffalini.github.io/slides/

PREMISE 1:

Neurodevelopmental disorders are (mostly) dimensional

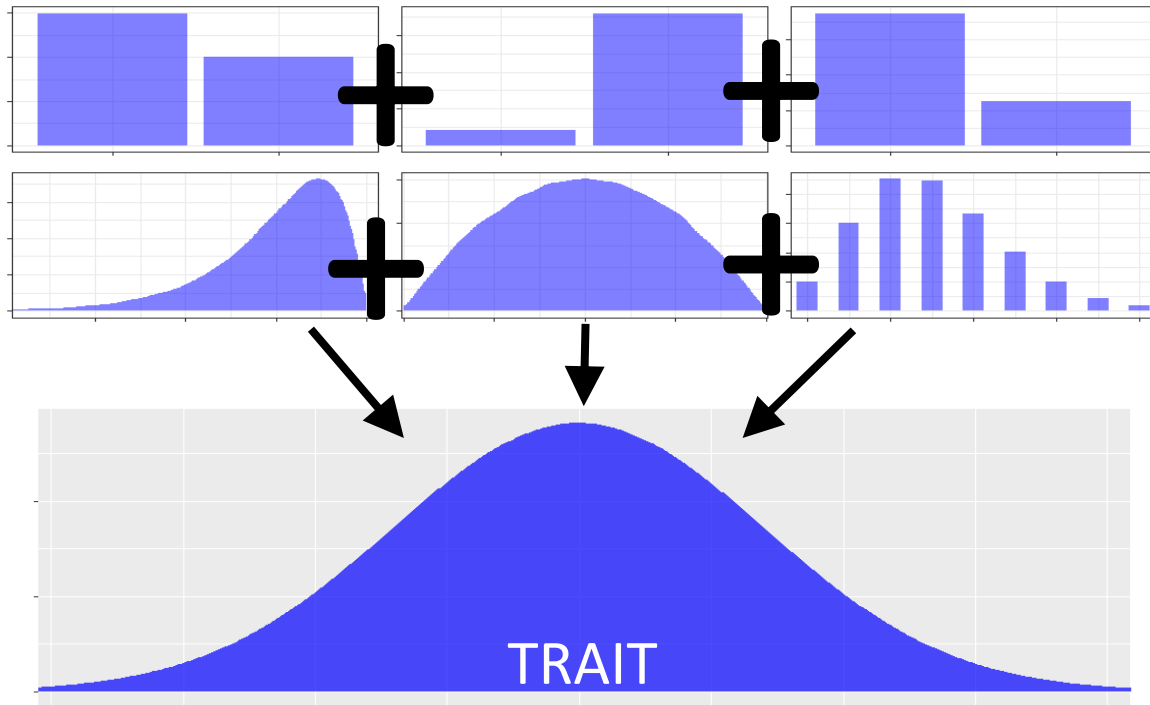
Although I will compare “clinical cases” and “controls”, these labels should **NOT** be mistaken for **natural categories**. For most conditions, diagnosis identifies **the extreme of continuously distributed traits**, and the cutoff is mainly a pragmatic clinical decision point.

*“Essentially **all behaviorally defined disorders** (...) are **continuous disorders**. Dyslexia is mainly defined as the low end of a **normal distribution** of word reading ability (Rodgers 1983; Shaywitz et al. 1992)”*

Peterson & Pennington (2015)

Why should neurodevelopmental traits be **continuous** – and even **normally distributed**?

Human behaviour reflects the cumulative influence of many small, partly independent **factors**—e.g., genetic polygenicity even in neurodevelopmental conditions (Plomin & Kovas, [2005](#); Kovas & Plomin, [2006](#); Grove et al., [2019](#); Demontis et al., [2023](#); Doust et al., [2022](#))



CENTRAL LIMIT THEOREM:
the sum of many independent variables, however distributed, tends to **converge** to a continuous *Gaussian* distribution

Diagnostic groups are imperfect approximations — do not reflect natural categories

Recent proposals aim to incorporate neurodevelopmental conditions within hierarchical dimensional frameworks of general psychopathology such as HiTOP (the genetic level, *polygenicity + pleiotropy* create a strong basis for a *hierarchical dimensional continuum*)

World Psychiatry
OFFICIAL JOURNAL OF THE WORLD PSYCHIATRIC ASSOCIATION (WPA)

2024

Special Article |  Free Access

Where do neurodevelopmental conditions fit in transdiagnostic psychiatric frameworks? Incorporating a new neurodevelopmental spectrum

Giorgia Michelini, Christina O. Carlisi, Nicholas R. Eaton, Jed T. Elison, John D. Haltigan, Roman Kotov, Robert F. Krueger, Robert D. Latzman, James J. Li ... [See all authors](#) ▾

First published: 16 September 2024 | <https://doi.org/10.1002/wps.21225> | Citations: 11

PREMISE 2:

It's (mostly) about g

In multicomponential batteries, a single *general factor (g)* **explains most common variance** among subtests and is **often the only psychometrically interpretable component**. The "specific" indices (visuospatial, working memory, etc.) have little reliability once "g" is considered

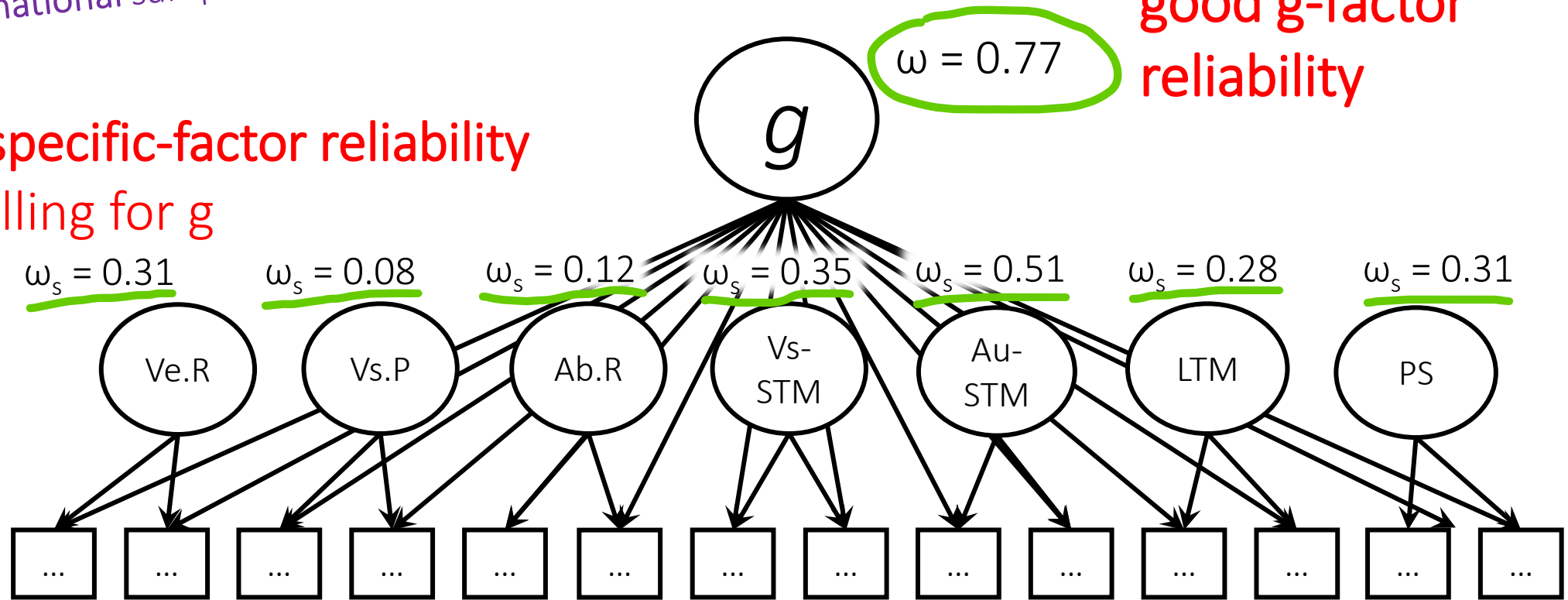
cf. Beaujean (2017); Canivez et al. (2016; 2017); Nelson et al., (2013); Watkins & Canivez (2022); Watkins (2010)

Factor-score **reliability** analysis using a bifactor model in the **Italian** standardization sample (N = 1281)

results closely mirror those obtained with the Wechsler scales and in international samples (reassuring... and not so reassuring 😊)

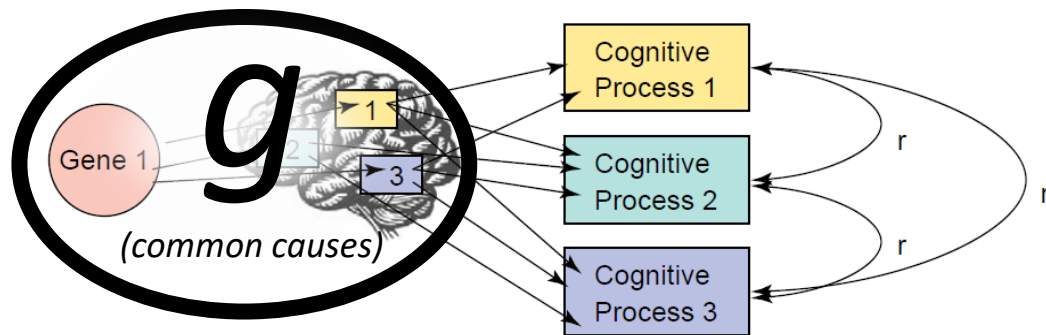
small to nil specific-factor reliability after controlling for g

good g-factor reliability



RMSEA = 0.04, SRMR = 0.03, CFI = 0.96, NNFI = 0.95

g is NOT just a statistical artifact



Pleiotropy: same genes influence different areas of the brain, that influences various cognitive processes

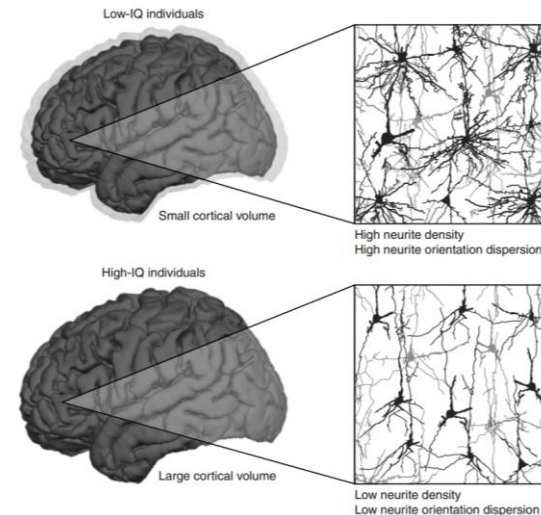


positive manifold

Pleiotropy “IS” g (common causes)

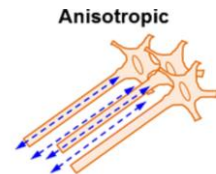
“Generalist” neurobiological bases

e.g., Haier (2017); Warne (2020)

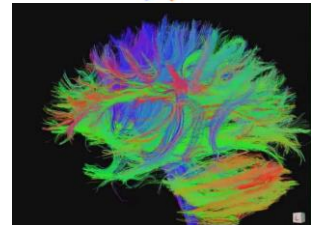


Total brain volume

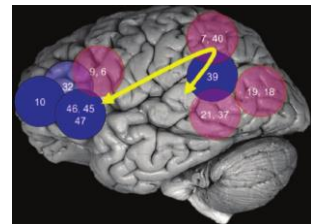
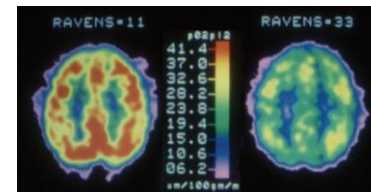
Efficient pruning and dendritic arborization



Integrity of fronto-parietal white-matter tracts



Efficient activation and switching during and across tasks



measuring g , in valid
and accurate ways, is
highly important

yet, in *neurodevelopmental conditions*, considering the distinctive patterns of strengths and weaknesses, alongside *g*, is still important!

(e.g., Cornoldi et al., 2015; Giofrè et al., 2017; 2019; Toffalini et al., 2017a, 2017b, 2022)

Strengths and Weaknesses in the Intellectual Profile of Different Subtypes of Specific Learning Disorder: A Study on 1,049 Diagnosed Children

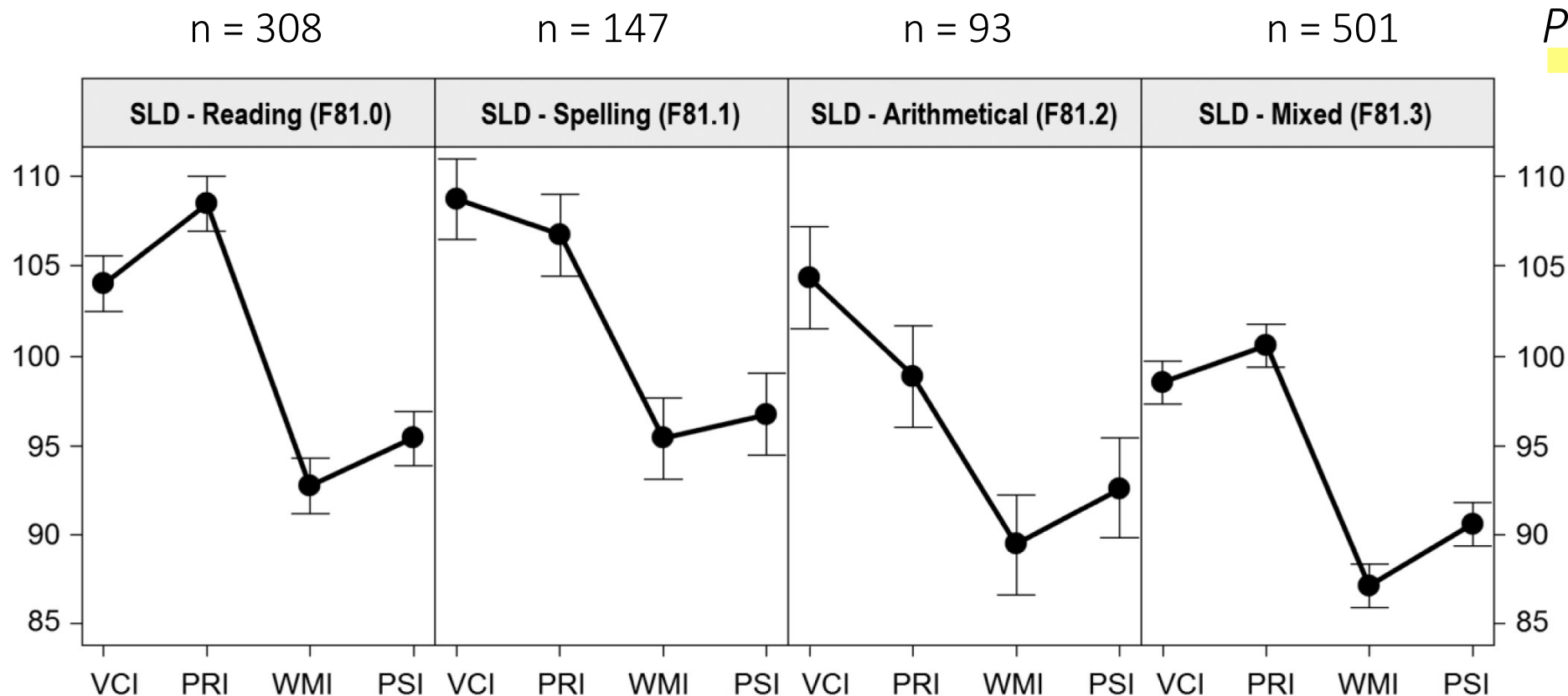
[Enrico Toffalini](#) ✉, [David Giofrè](#), and [Cesare Cornoldi](#) [View all authors and affiliations](#)

[Volume 5, Issue 2](#) | <https://doi.org/10.1177/2167702616672038>

Large LD sample ($N = 1049$, $k = 27$) collected in Italy via the AIRIPA network

- ~10-15 IQ points between *General Ability* areas and basic *Cognitive Processing* areas

- Minor differences across LD subtypes

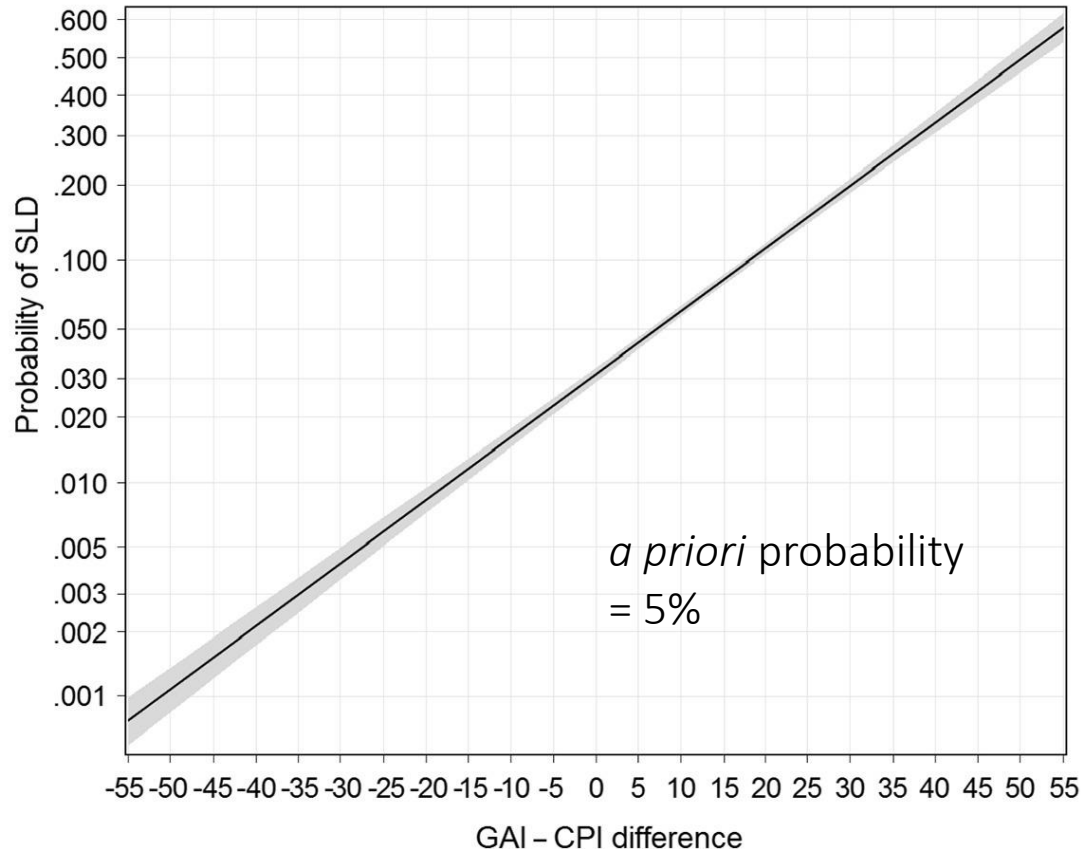


Intelligence measures as diagnostic tools for children with specific learning disabilities

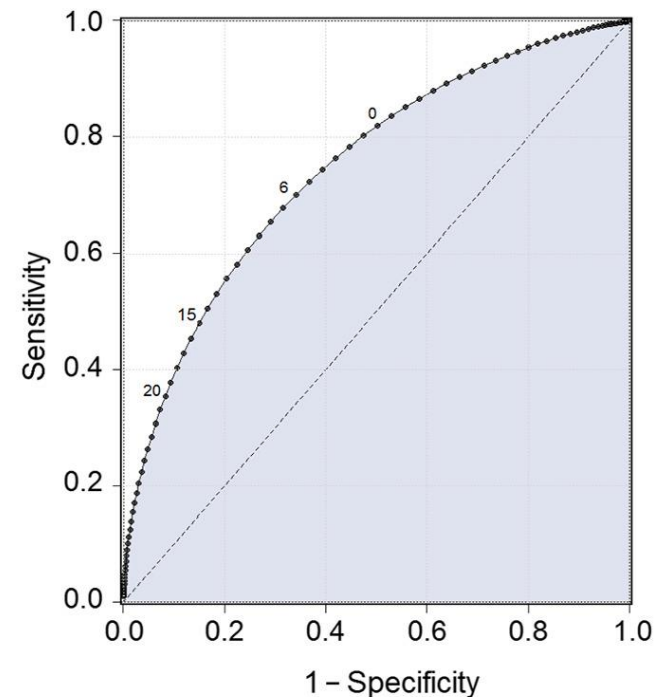
David Giofrè^a, Enrico Toffalini^b, Gianmarco Altoè^c, Cesare Cornoldi^b

- Average internal discrepancy or profile enough to significantly affect the propensity to give a diagnosis

A) Probability plot



B) ROC curve for the GAI-CPI difference



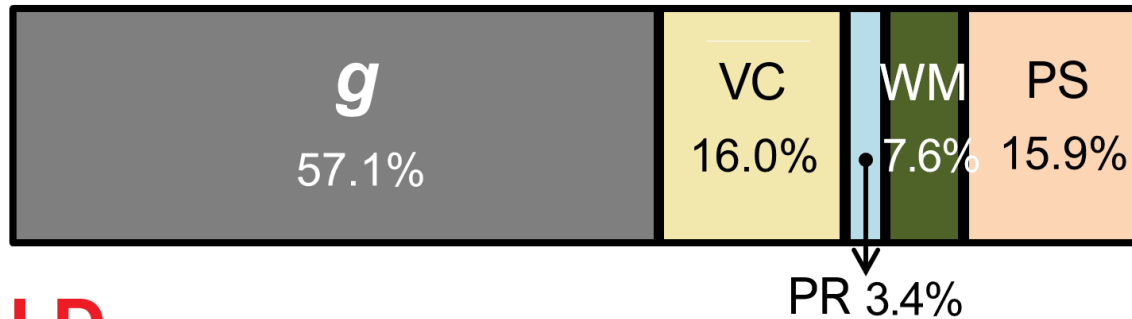
- Discriminates LD vs Neurotypical; *NOT* across LD subtypes

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

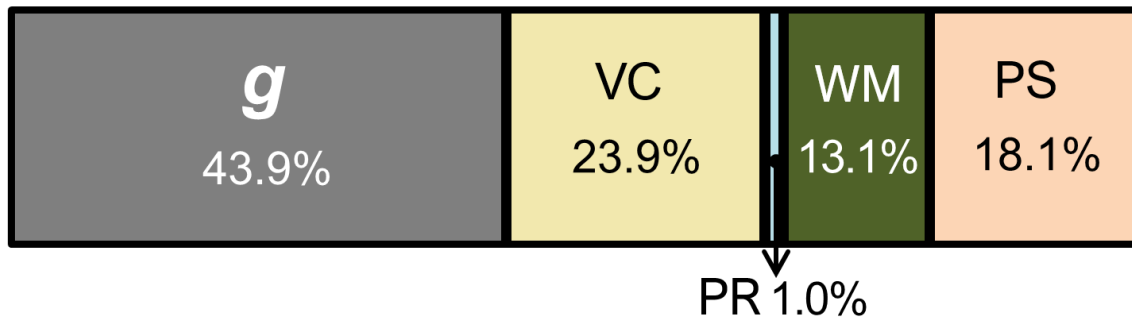
David Giofrè ^a  , Massimiliano Pastore ^b, Cesare Cornoldi ^c, Enrico Toffalini ^c  

subtest common variance explained in...

Typical dev. (IQ ≥ 70)



LD



- g explains 57% of common variance in TD but only 44% in LD

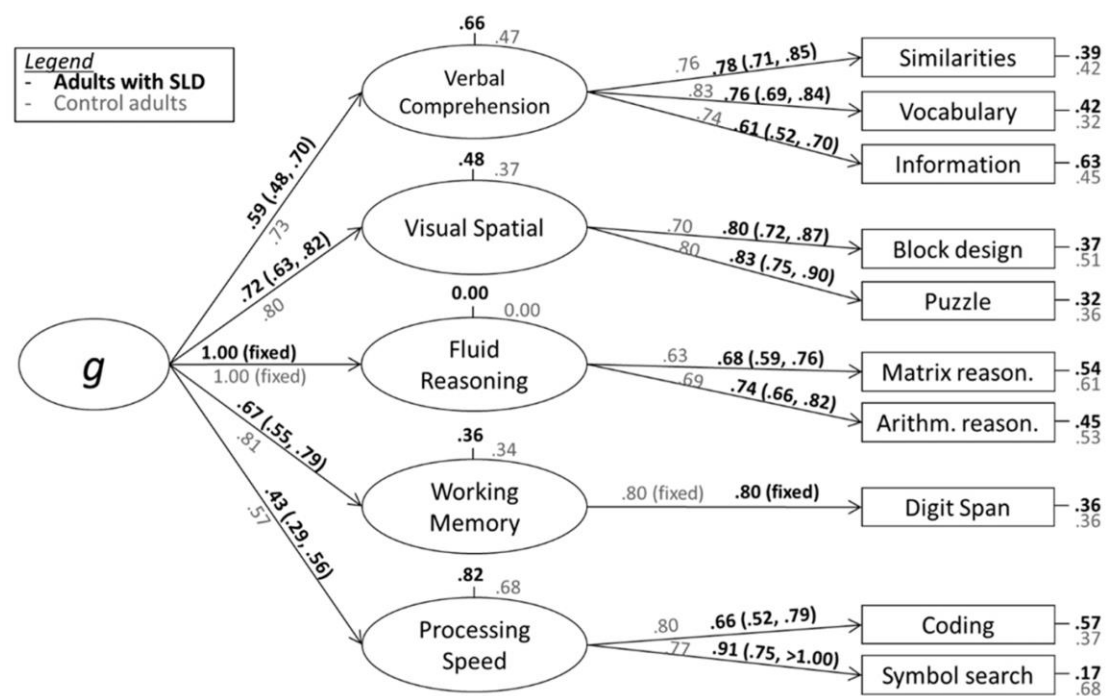
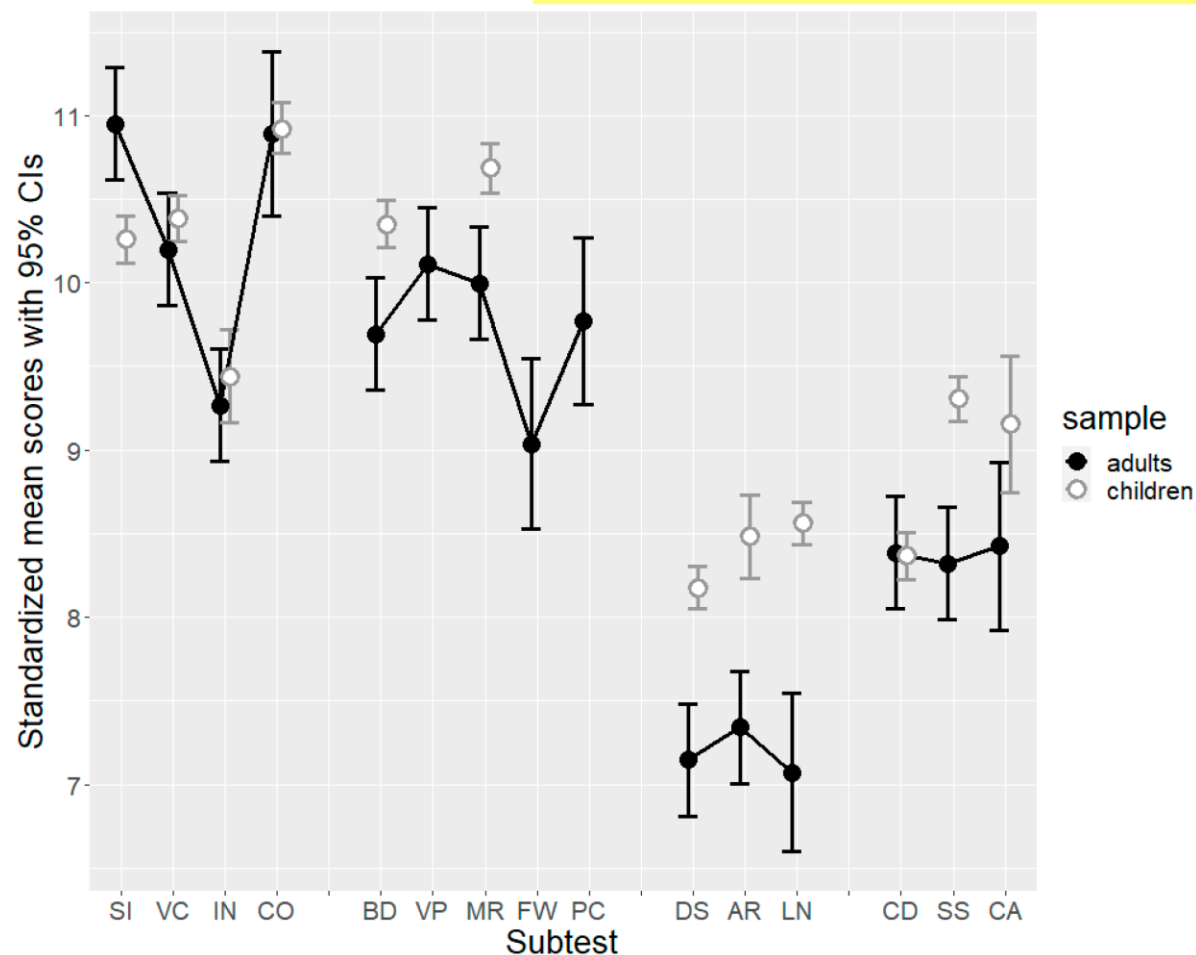
- there is generally more room for interpretation of domain-specific factor scores in LD than in TD

- however, reliability estimates (ω) remain too limited for independent interpretation even in LD (< 0.60)

The Intellectual Profile of Adults with Specific Learning Disabilities

by Eleonora Pizzigallo ^{1,*} , Cesare Cornoldi ¹, Serafino Buono ², Santina Città ², Francesco Viola ¹ and Enrico Toffalini ¹ 

Quantitatively similar findings also in adult population (N = 301)

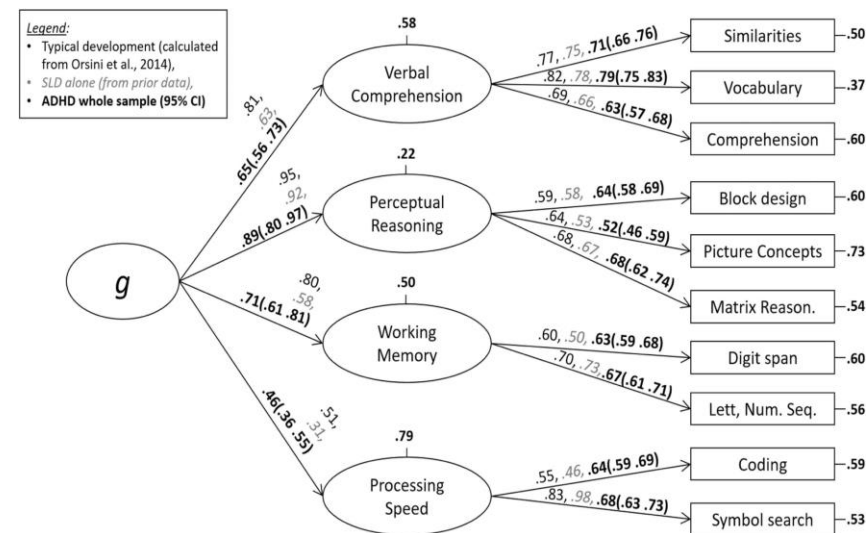
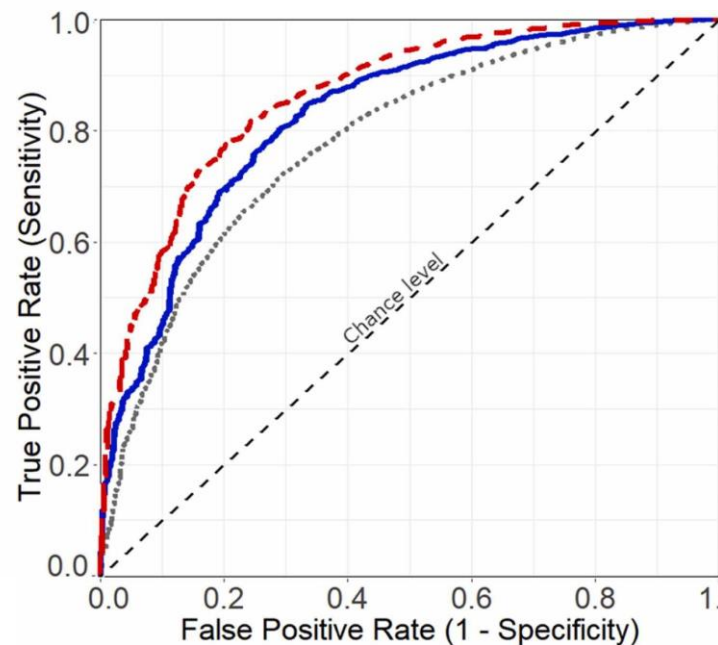
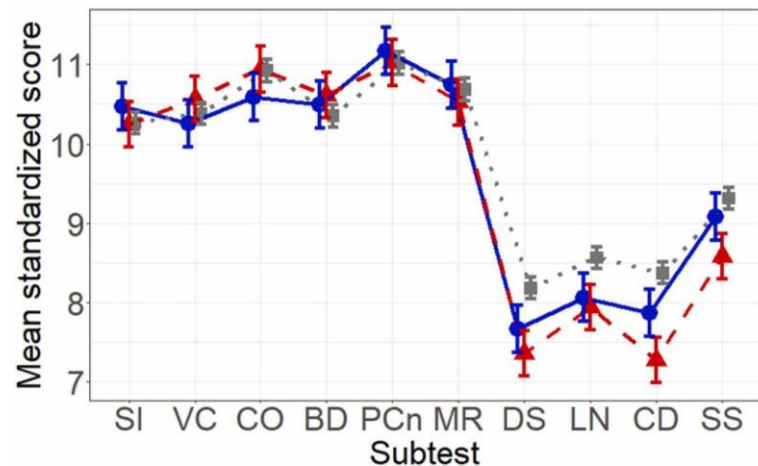
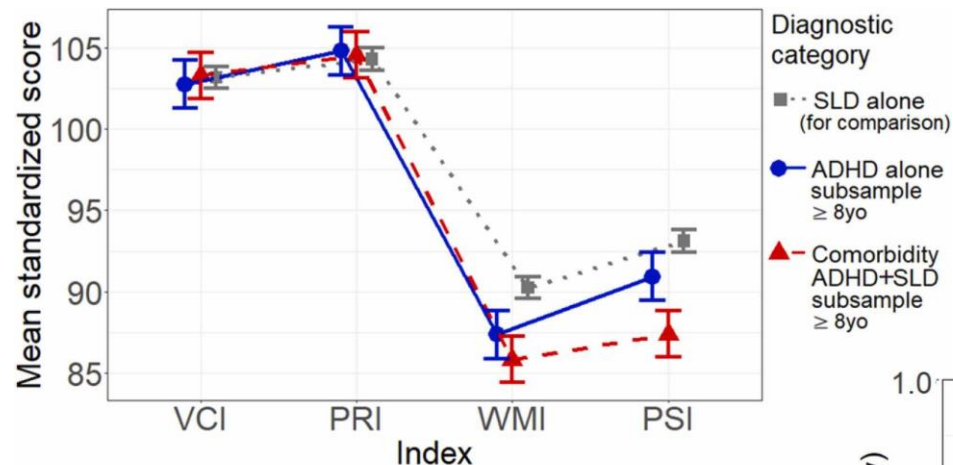


The structure, profile, and diagnostic significance of intelligence in children with ADHD are impressively similar to those of children with a specific learning disorder

Enrico Toffalini ^{a 1}, Serafino Buono ^{b 2}, Cesare Cornoldi ^{a 3} [Show more](#)

Essentially same conclusions in ADHD (N = 948)

- Meaningful differences are found when contrasting *Neurodev.Cond.* vs *TD*, not so much across categories of *Neurodev.Cond.s*





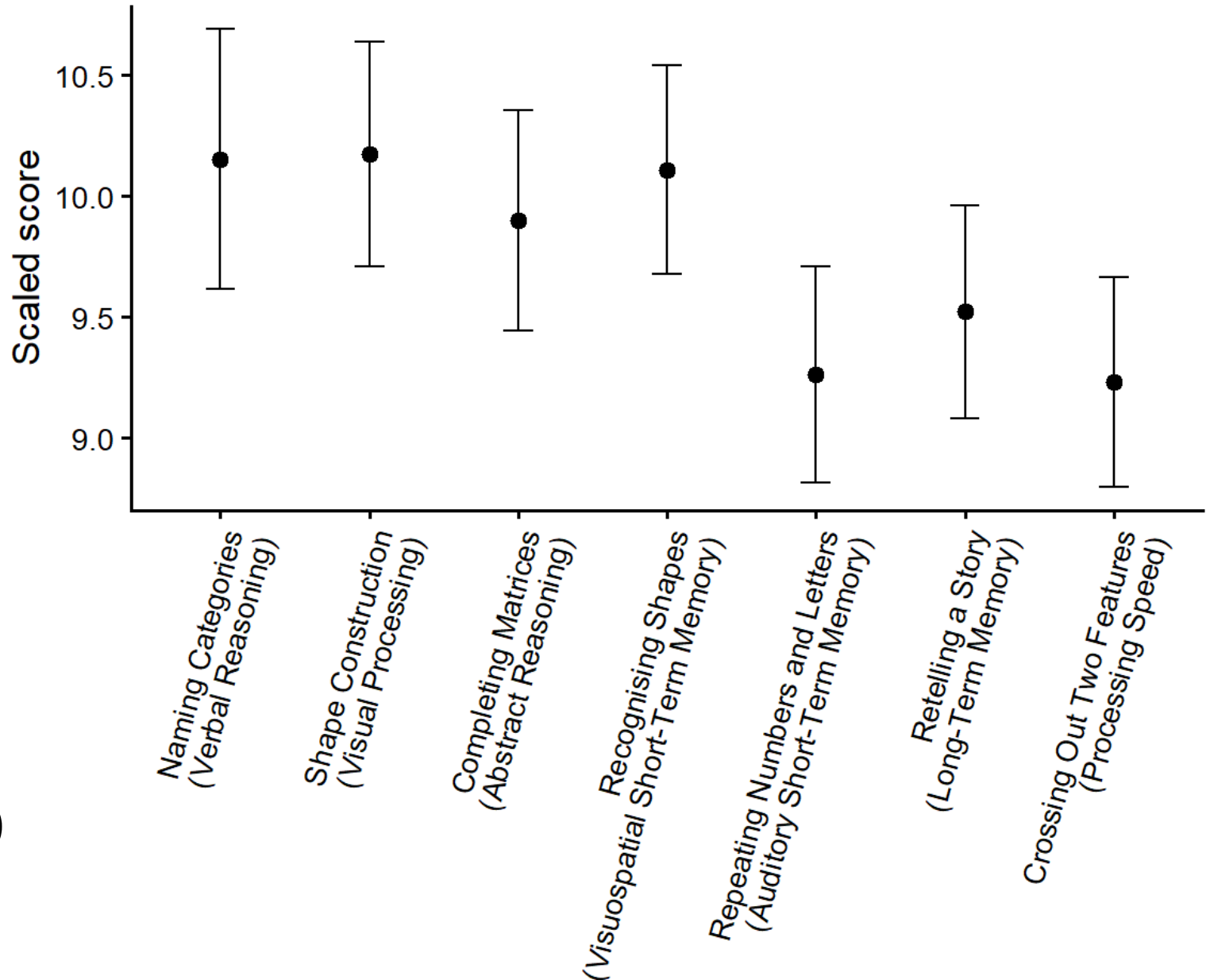
Preliminary analysis on clinical cases with IDS-2: Average profiles of subtests

N = 138, k = 4

only Italian clinical cases

- The well-known average profile!
General Ability areas > Cognitive
Proficiency areas ... but with
smaller discrepancy ? (<1 SS point)*

(LD: N = 112; ADHD: N = 19; both: N = 7)





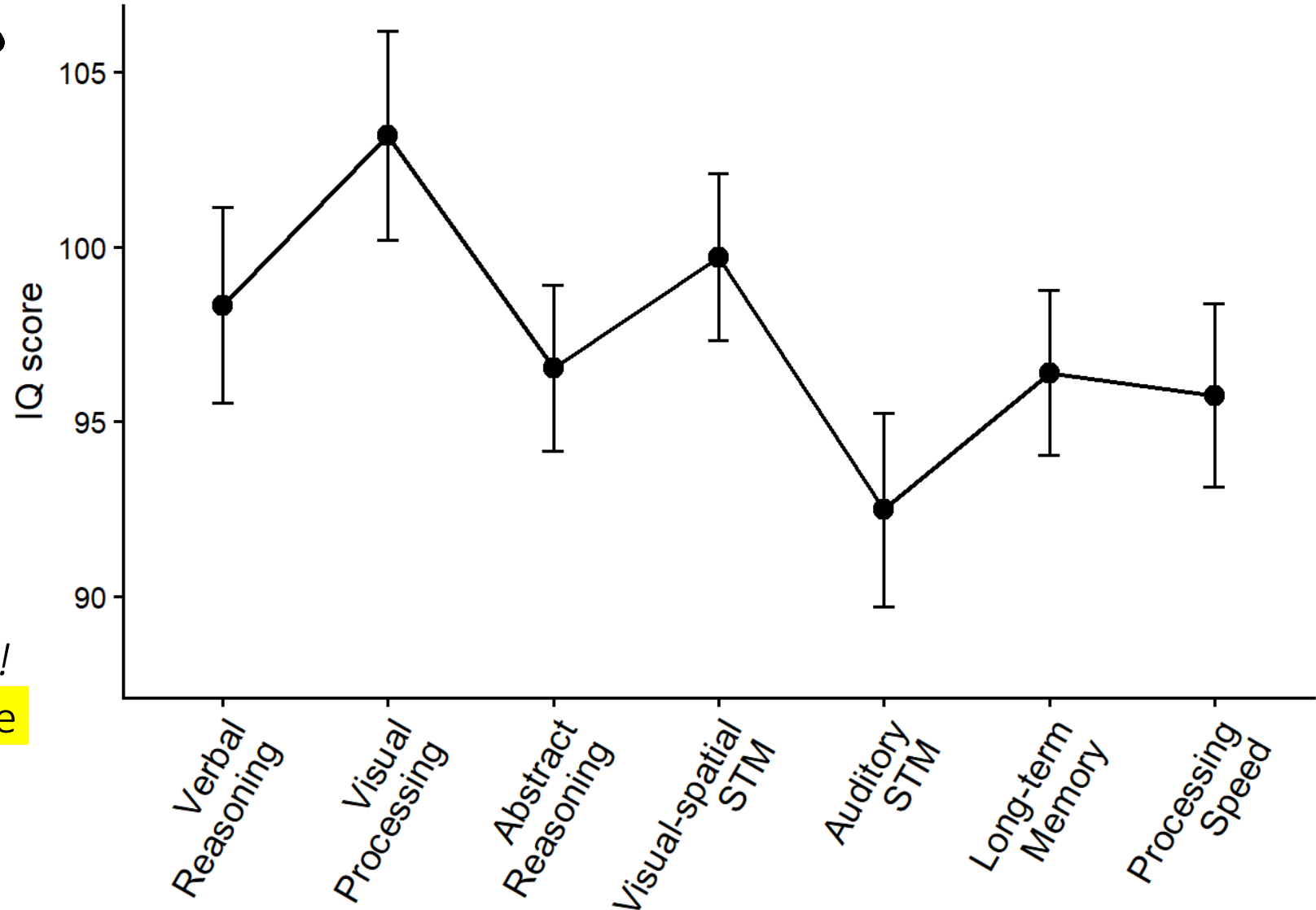
Preliminary analysis on clinical cases with IDS-2: Average profiles of indices

N = 118, k = 4

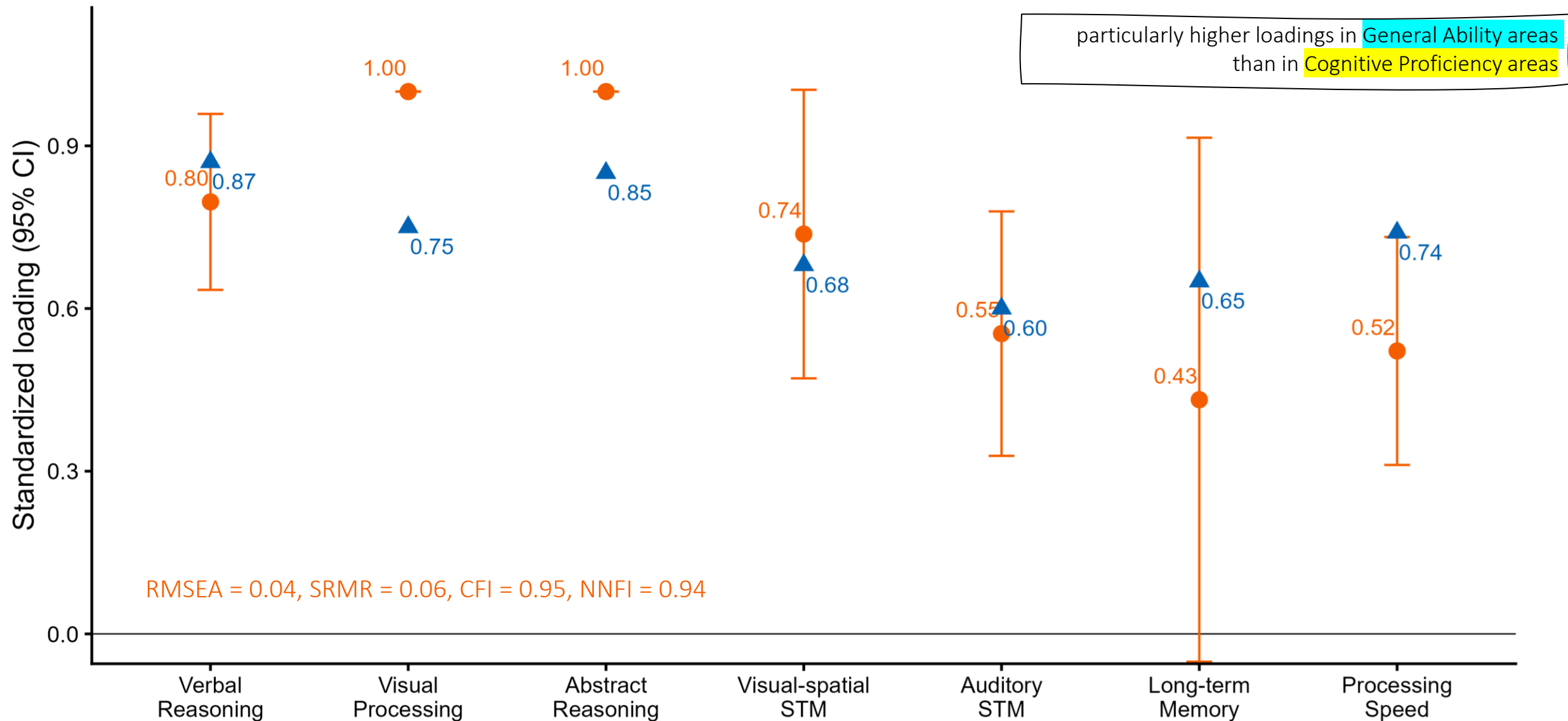
only Italian clinical cases

- *The well-known average profile!
General Ability areas > Cognitive
Proficiency areas ... but with
smaller discrepancy? (~5-10 IQ
points)*

(LD: N = 100; ADHD: N = 14; both: N = 4)



Color: **blue** for standardization sample; **orange** for clinical sample



“Central Ability Index” (CAI)
a better way to measure *g* in
LD, ADHD, etc. (?)

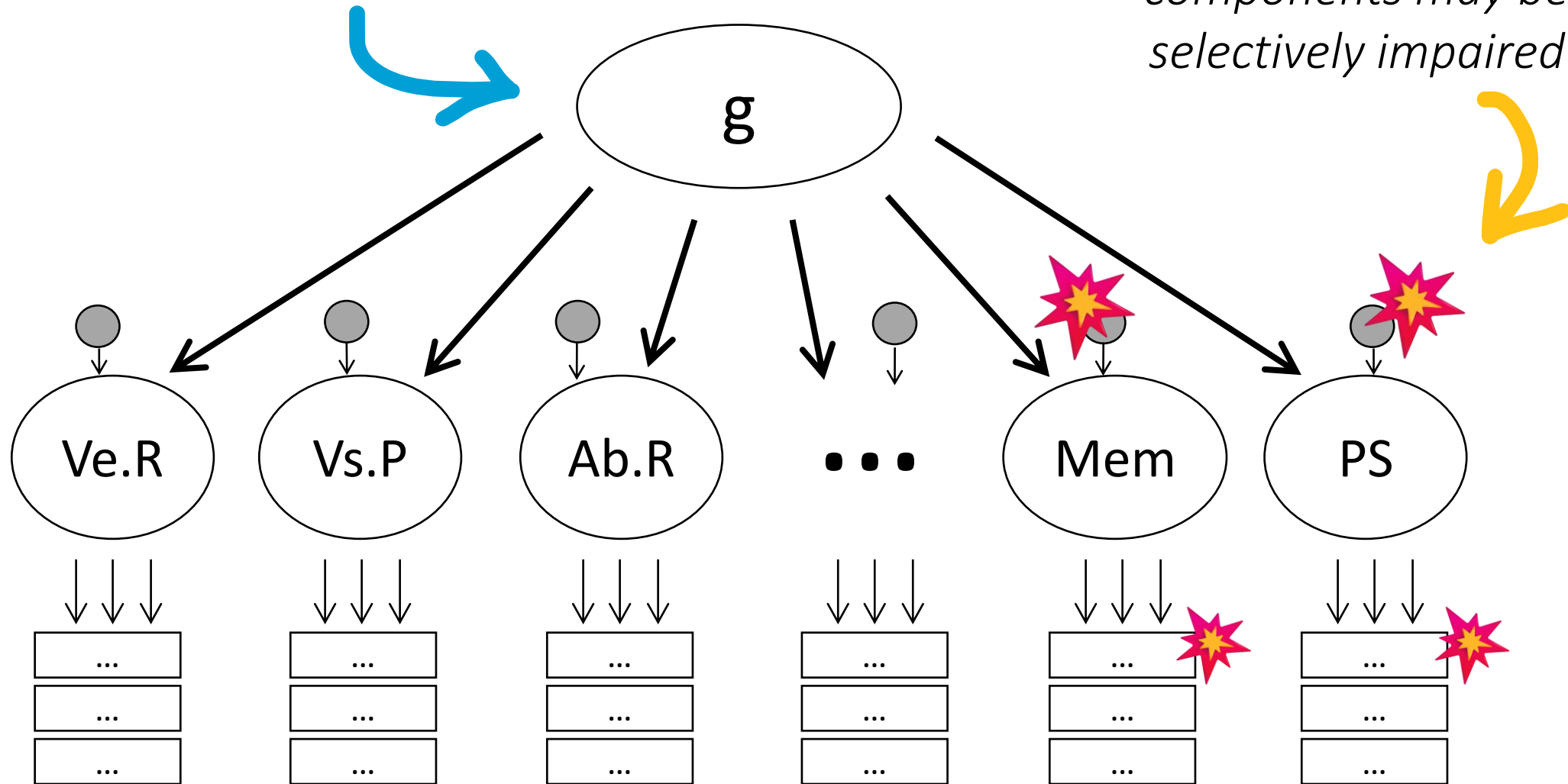
(analogous to the *General Ability Index*
from the Wechsler scales)

Why is CAI needed?

- Our research & previous literature (e.g., Raiford et al., 2005; Saklofske et al., 2005), suggest that areas involving attention + basic cognitive proficiency (WM, PS) may be selectively impaired in LD, ADHD, and other neuropsychological conditions
 - WM and PS are less related to g than other areas (e.g., Giofrè et al., 2017)
 - Condition-related impairments are already clinically considered
- Better keep WM and PS apart from g -factor estimate in clinical cases

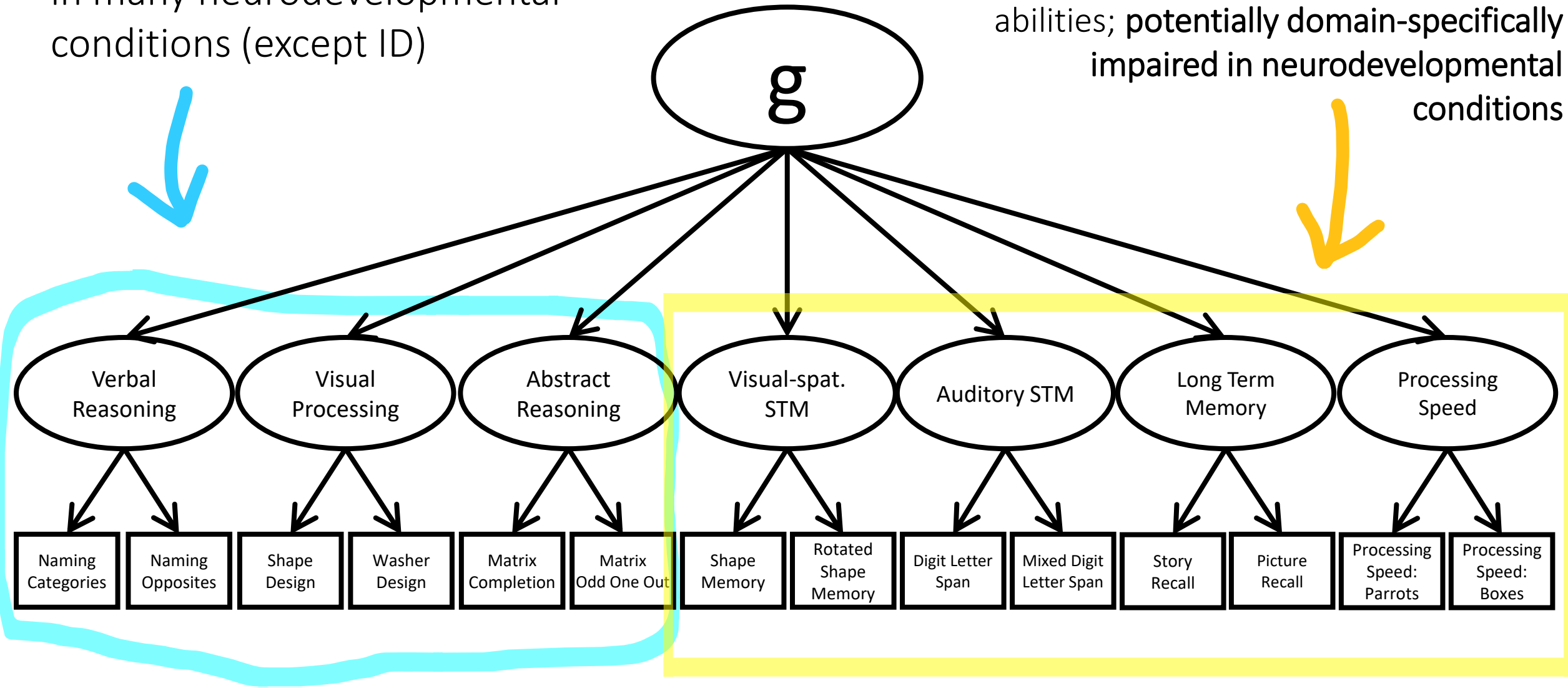
g **not** necessarily impaired (at least after discounting the clinically-relevant impairment)

some domain-specific components may be selectively impaired



areas **not** necessarily impaired
in many neurodevelopmental
conditions (except ID)

involve basic cognitive processing,
attention, memory, phonological
abilities; **potentially domain-specifically**
impaired in neurodevelopmental
conditions



Proposal for a CAI index

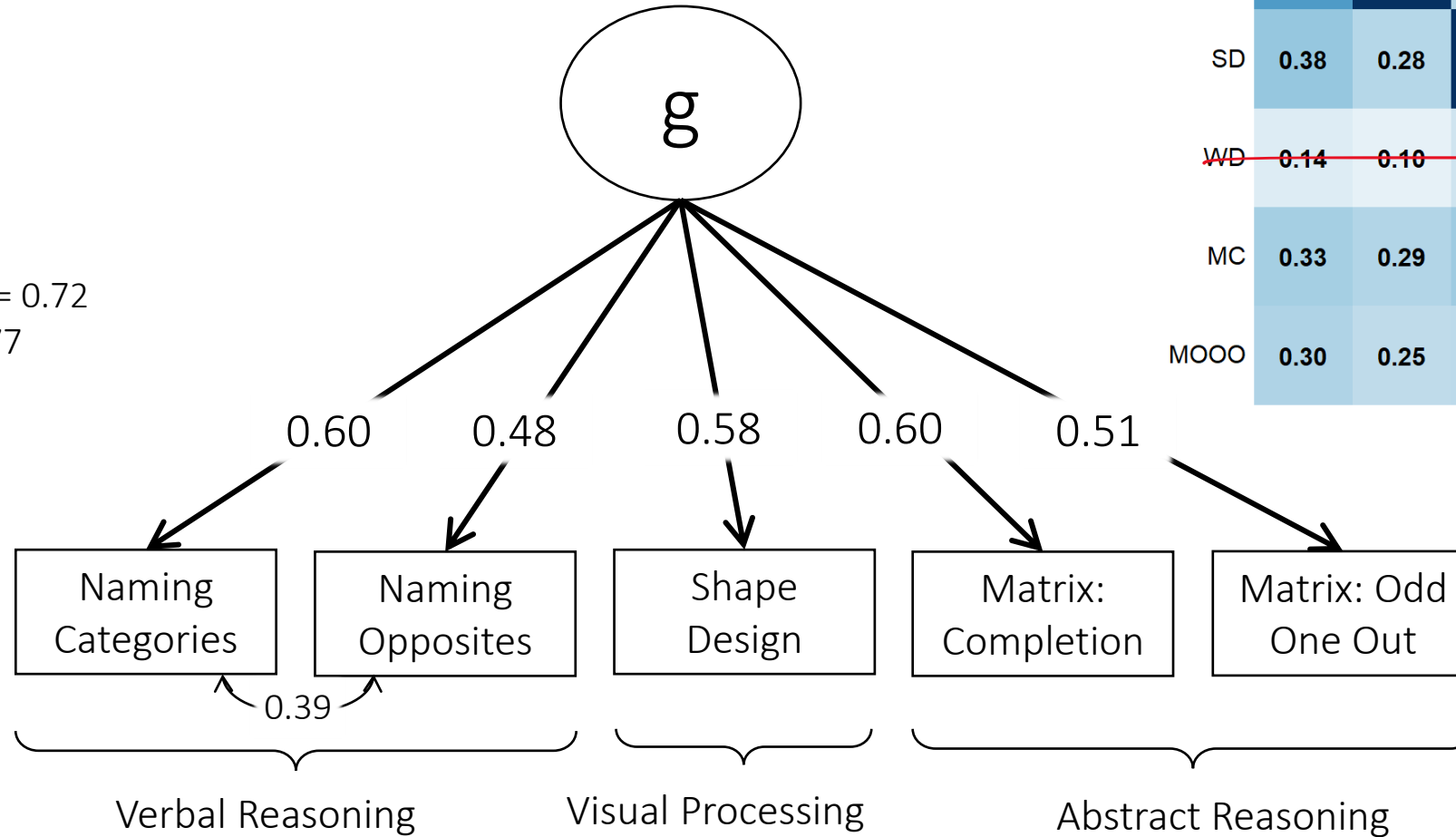
Reliability

CAI: $\alpha = 0.71$, $\omega = 0.66$

screening IQ: $\alpha = 0.50$

general IQ: $\alpha = 0.72$; $\omega = 0.72$

full IQ: $\alpha = 0.82$; $\omega = 0.77$



	NCa	NO	SD	WD	MC	MOOO
NCa	1.00	0.56	0.38	0.14	0.33	0.30
NO	0.56	1.00	0.28	0.10	0.29	0.25
SD	0.38	0.28	1.00	0.20	0.35	0.26
WD	0.14	0.10	0.20	1.00	0.19	0.10
MC	0.33	0.29	0.35	0.19	1.00	0.34
MOOO	0.30	0.25	0.26	0.10	0.34	1.00

RMSEA = 0.04, SRMR = 0.02, CFI = 0.99, NNFI = 0.98

CAI

Conversion table

from Scale Score sum (PP) to CAI based on su
5 subtest in IDS-2 (analytical computation)

Toffalini, E., & Cornoldi, C. (2025). *Calcolo di un
indice di abilità centrale (IAC) per la batteria di
intelligenza IDS-2* [Preprint]. PsyArXiv.
<https://doi.org/10.31234/osf.io/b5rh8> v1

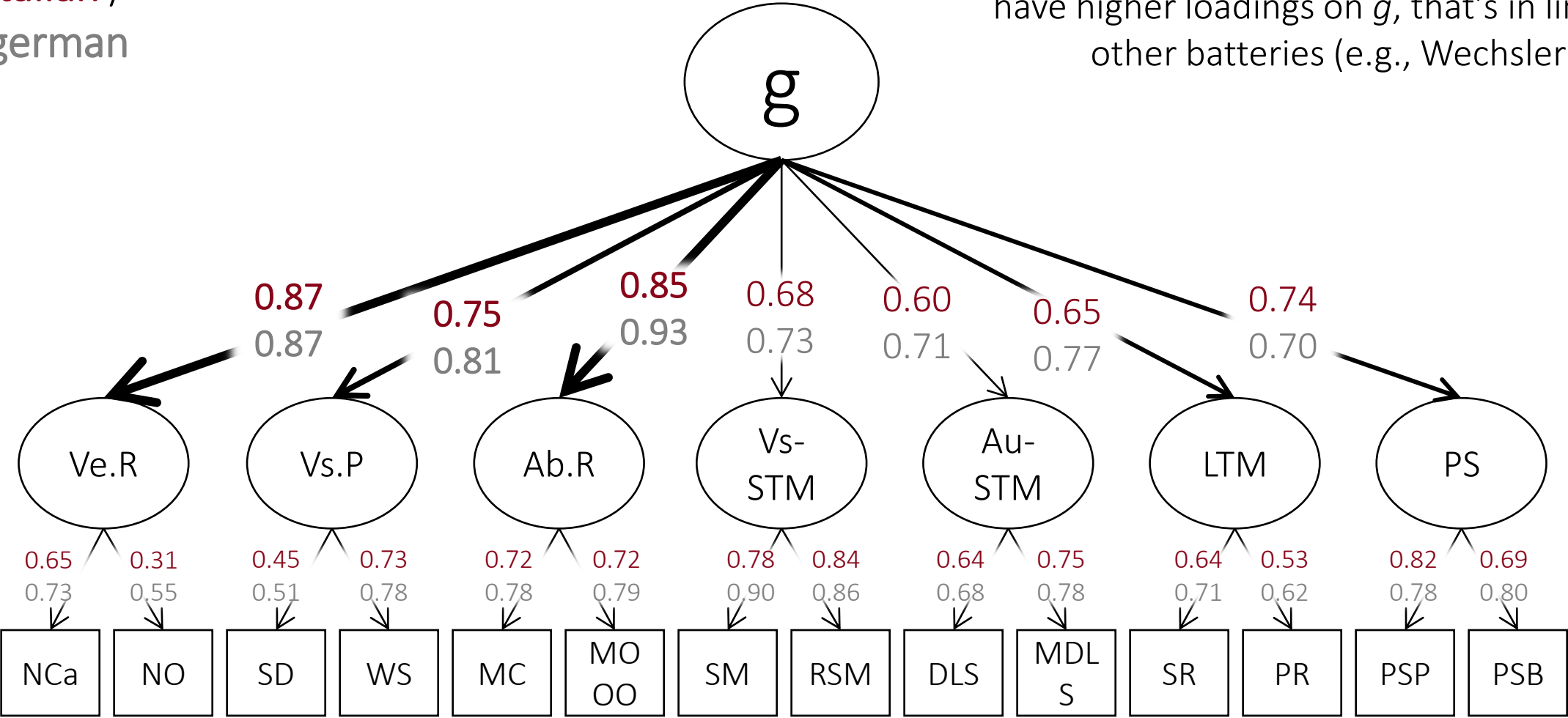
PP	IAC	PP	IAC	PP	IAC	PP	IAC	PP	IAC
5	34	24	62	43	90	62	118	81	145
6	36	25	63	44	91	63	119	82	147
7	37	26	65	45	93	64	120	83	148
8	39	27	66	46	94	65	122	84	150
9	40	28	68	47	96	66	123	85	151
10	41	29	69	48	97	67	125	86	153
11	43	30	71	49	99	68	126	87	154
12	44	31	72	50	100	69	128	88	156
13	46	32	74	51	101	70	129	89	157
14	47	33	75	52	103	71	131	90	159
15	49	34	77	53	104	72	132	91	160
16	50	35	78	54	106	73	134	92	161
17	52	36	80	55	107	74	135	93	163
18	53	37	81	56	109	75	137	94	164
19	55	38	82	57	110	76	138	95	166
20	56	39	84	58	112	77	140		
21	58	40	85	59	113	78	141		
22	59	41	87	60	115	79	142		
23	60	42	88	61	116	80	144		

“Balanced IQ” (BIQ)

Perhaps a better way to
measure *g*-factor ... in general

CFA for standardization samples:
italian /
german

«Verbal Reasoning» (crystallized), «Visual Processing», and «Abstract Reasoning» (fluid) have higher loadings on *g*, that's in line with other batteries (e.g., Wechsler scales)



- CHC **never** made us believe that *broad abilities* must have the exact same loadings on *g* → different loadings are expected!
- clinicians will NOT weight subtests by loadings when computing the “factor scores” for *g* (IQ) ...
- WISC-IV settled this by including different number of subtests for different areas (e.g., 3 for Perceptual Reasoning, 2 for Working Memory)
- More broadly, IDS-2 samples a bit too many subtests from basic cognitive proficiency areas (Memory + Speed = 4/7 of the battery)

Proposal for a BIQ index

Reliability

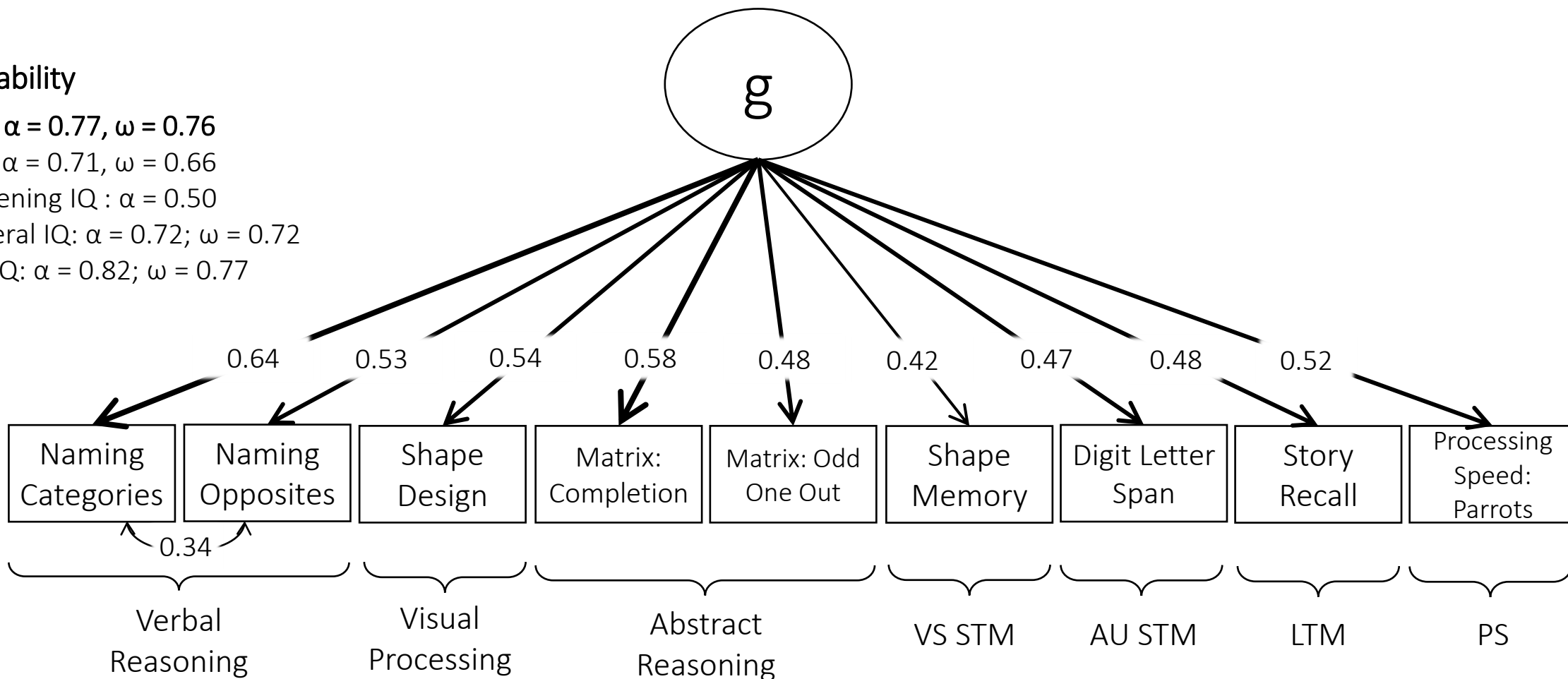
BIQ: $\alpha = 0.77$, $\omega = 0.76$

CAI: $\alpha = 0.71$, $\omega = 0.66$

screening IQ : $\alpha = 0.50$

general IQ: $\alpha = 0.72$; $\omega = 0.72$

full IQ: $\alpha = 0.82$; $\omega = 0.77$



RMSEA = 0.05, SRMR = 0.03, CFI = 0.96, NNFI = 0.94

BIQ

Conversion Table

from Scaled Scores sum (PP) to BIQ based on
9 subtests in IDS-2 (analytical computation)

details available via OSF:

<https://doi.org/10.17605/OSF.IO/SRBG5>

PP	QIB	PP	QIB	PP	QIB	PP	QIB
9	24	50	63	91	101	132	139
10	25	51	64	92	102	133	140
11	26	52	65	93	103	134	141
12	27	53	65	94	104	135	142
13	28	54	66	95	105	136	143
14	29	55	67	96	106	137	144
15	30	56	68	97	107	138	145
16	31	57	69	98	107	139	146
17	32	58	70	99	108	140	147
18	33	59	71	100	109	141	148
19	34	60	72	101	110	142	148
20	35	61	73	102	111	143	149
21	36	62	74	103	112	144	150
22	37	63	75	104	113	145	151
23	38	64	76	105	114	146	152
24	38	65	77	106	115	147	153
25	39	66	78	107	116	148	154
26	40	67	79	108	117	149	155
27	41	68	79	109	118	150	156
28	42	69	80	110	119	151	157
29	43	70	81	111	120	152	158
30	44	71	82	112	121	153	159
31	45	72	83	113	121	154	160
32	46	73	84	114	122	155	161
33	47	74	85	115	123	156	162
34	48	75	86	116	124	157	162
35	49	76	87	117	125	158	163
36	50	77	88	118	126	159	164
37	51	78	89	119	127	160	165
38	52	79	90	120	128	161	166
39	52	80	91	121	129	162	167
40	53	81	92	122	130	163	168
41	54	82	93	123	131	164	169
42	55	83	93	124	132	165	170
43	56	84	94	125	133	166	171
44	57	85	95	126	134	167	172
45	58	86	96	127	135	168	173
46	59	87	97	128	135	169	174
47	60	88	98	129	136	170	175
48	61	89	99	130	137	171	176
49	62	90	100	131	138		

Really the last slide

- CAI and BIQ can be computed from existing IDS-2 subtests, and their internal structures are psychometrically plausible
- Reasons for using them are theory-driven  CAI: knowledge of peculiarities of LD, ADHD
 BIQ: better reflecting CHC and loadings / factor-scores
- Convergent validity with other batteries (e.g., Wechsler)? To be established
- Measurement invariance across countries, gender, clinical conditions, etc.? To be established

enricotoffalini.github.io/slides/

