

1
2
3
4
5
6
7
8
9
10
11

Title:

A hierarchical watershed model of fluid intelligence in childhood and adolescence

Running Title:

Fluid intelligence in childhood and adolescence

Authors:

Fuhrmann, D.^{1*}, Simpson-Kent, I. L.¹, Bathelt, J.¹, the CALM team¹ & Kievit, R. A.¹

¹ MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, CB2 7EF, UK

* corresponding author: Delia Fuhrmann, MRC CBU, 15 Chaucer Road, Cambridge, CB2 7EF,

UK, email: delia.fuhrmann@mrc-cbu.cam.ac.uk

12

Abstract

13 Fluid intelligence is the capacity to solve novel problems in the absence of task-specific
14 knowledge, and is highly predictive of outcomes like educational attainment and
15 psychopathology. Here, we modelled the neurocognitive architecture of fluid intelligence in
16 two cohorts: CALM ($N = 551$, aged 5 - 17 years) and NKI-RS ($N = 335$, aged 6 - 17 years). We
17 used multivariate Structural Equation Modelling to test a preregistered watershed model of
18 fluid intelligence. This model predicts that white matter contributes to intermediate cognitive
19 phenotypes, like working memory and processing speed, which, in turn, contribute to fluid
20 intelligence. We found that this model performed well for both samples and explained large
21 amounts of variance in fluid intelligence ($R^2_{\text{CALM}} = 51.2\%$, $R^2_{\text{NKI-RS}} = 78.3\%$). The relationship
22 between cognitive abilities and white matter differed with age, showing a dip in strength
23 around ages 7 - 12 years. This age-effect may reflect a reorganization of the neurocognitive
24 architecture around pre- and early puberty. Overall, these findings highlight that intelligence
25 is part of a complex hierarchical system of partially independent effects.

26

Keywords

27 Working memory, processing speed, fractional anisotropy, watershed model, structural
28 equation modeling

29 Fluid intelligence (g_f) is a core part of human cognition and refers to the capacity to solve
30 novel problems in the absence of task-specific knowledge. It is highly predictive of a number
31 of important life span outcomes, including educational attainment (Primi et al. 2010; Roth et
32 al. 2015) and psychopathology (Gale et al. 2010). Despite years of investigation, however, our
33 understanding of the neurocognitive architecture of g_f remains limited. Longstanding debates
34 have considered, for instance, how g_f relates to more fundamental cognitive functions such
35 as working memory and processing speed, and how all of these cognitive functions relate to
36 brain structure and function (Kyllonen and Christal 1990; Fry and Hale 2000; Chuderski 2013;
37 Ferrer et al. 2013).

38 Working memory is the ability to hold and manipulate information in the mind short-term. It
39 has been suggested that working memory is a key determinant of g_f by limiting mental
40 information processing capacity (Fukuda et al. 2010; Chuderski 2013). Proponents of this
41 working memory account of g_f cite high correlations between the two domains ranging from
42 0.5 to 0.9 in meta-analyses (Ackerman et al. 2005; Oberauer et al. 2005). Such high
43 correlations have led some to suggest that g_f and working memory are, in fact, isomorphic
44 (Kyllonen and Christal 1990). However, more recent work has highlighted that this
45 isomorphism only arises under conditions of high time constraints for g_f tasks (Chuderski
46 2013). This suggests that g_f and working memory are, in fact, separable constructs and
47 underlines the importance of processing speed for g_f .

48 Processing speed, the speed of mental computations, is thought to be rate-limiting to g_f and
49 is therefore sometimes proposed to be a particularly good predictor of g_f (Kail and Salthouse
50 1994; Salthouse 1996; Ferrer et al. 2013; Kail et al. 2015; Schubert et al. 2017). Proponents of
51 the processing speed account of g_f cite moderate but robust correlations between g_f and

52 processing speed of 0.2 in meta-analyses (Sheppard and Vernon 2008) as well as longitudinal
53 evidence (Finkel et al. 2005; Coyle et al. 2011; Kail et al. 2015). Salthouse (1996) argued in the
54 context of cognitive aging, that processing speed determines high-level cognitive
55 performance because slow processing means that relevant sub-operations cannot be
56 completed in a set amount of time or are not available for successful integration. A
57 complementary explanation of individual differences in g_f proposes that processing speed
58 may be a direct reflection of fundamental neuroarchitectonic properties of the brain, such as
59 myelination or white matter microstructure (Lu et al. 2011; Chevalier et al. 2015).

60 White matter shows protracted development throughout childhood and adolescence, and
61 into the third decade of life (Mills et al. 2016). White matter tracts can be characterised *in*
62 *vivo* using diffusion-tensor imaging (DTI), which is sensitive, but not necessarily specific, to
63 white matter microstructural properties such as myelination or axonal density (Jones et al.
64 2013; Wandell 2016). Fractional anisotropy (FA) is the most commonly investigated DTI
65 measure and quantifies the directionality of water diffusion in different white matter tracts
66 (Pfefferbaum et al. 2000; Wandell 2016). Working memory, processing speed and g_f have
67 each been linked to individual differences in FA (Vestergaard et al. 2011; Kievit, Davis,
68 Griffiths, Correia, CamCAN, et al. 2016; Bathelt et al. 2018). While some studies, using
69 Principal Component Analysis, have posited that FA in different tracts can be summarized by
70 sizable single components (Penke et al. 2010; Cox et al. 2016), formal investigations using
71 confirmatory factor analysis have demonstrated that single-factor models of FA generally
72 show poor fit and do not adequately capture individual differences in white matter
73 microstructure (Lövdén et al. 2013; Kievit, Davis, Griffiths, Correia, Cam-CAN, et al. 2016). In
74 a similar vein, there is a growing body of literature showing specific associations between
75 white matter tracts and cognitive abilities, with those connecting frontoparietal regions

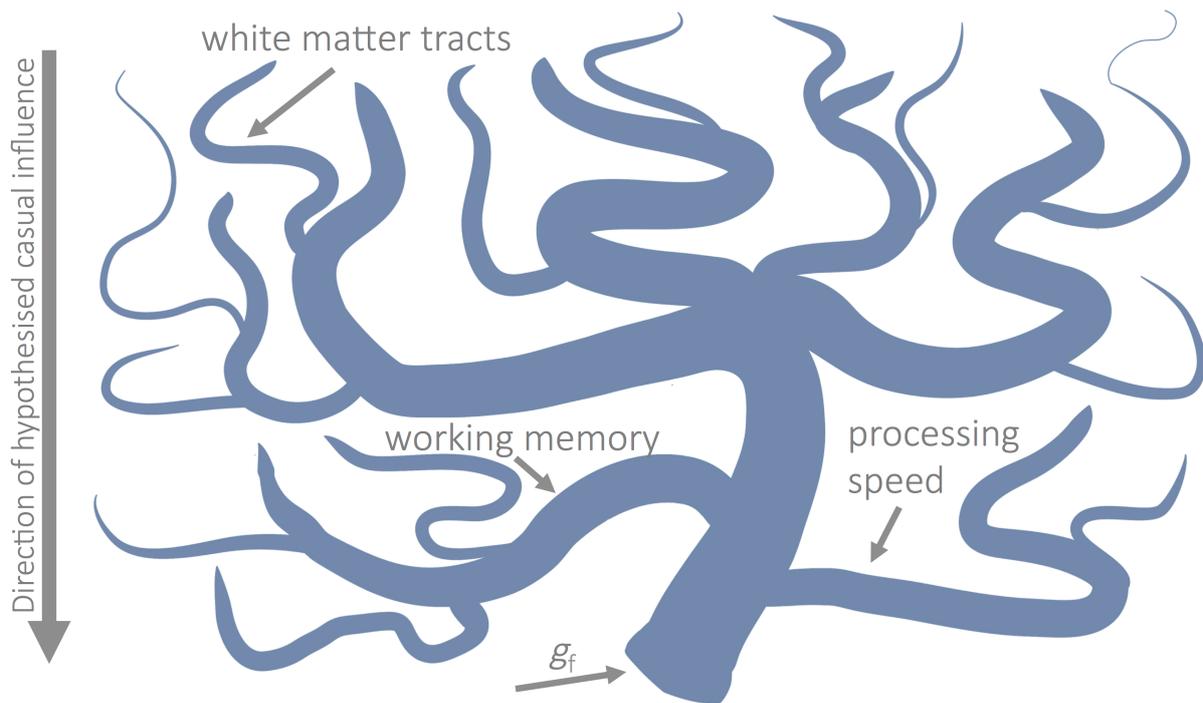
76 usually showing largest contributions to complex cognitive functions like g_f (Vestergaard et al.
77 2011; Kievit et al. 2016; Bathelt et al. 2018).

78 We here seek to address several critical outstanding issues in the field: First, there is limited
79 systematic evidence on the concurrent relationships between g_f , working memory,
80 processing speed and white matter. This leaves the relative contributions of processing speed
81 and working memory to g_f unclear, which, in turn, poses challenges for the design of effective
82 cognitive training interventions. Second, studies usually use a single task as a proxy for
83 complex and abstract constructs such as processing speed, working memory, and g_f . This
84 raises questions about the generalizability of findings (Noack et al. 2014). Third, our
85 understanding of how the relationships between relevant cognitive domains and between
86 brain and cognition change with age remains limited, raising the possibility that brain-
87 behaviour relationships may change with age (Garrett 1946; Johnson 2000; Tamnes et al.
88 2017).

89 To address these issues, we here used structural equation modelling (SEM) to model the
90 associations between g_f , working memory, processing speed, and white matter
91 microstructure and age in two large, independent samples: the Centre for Attention, Learning
92 and Memory sample (CALM, $N = 551$, aged 5 - 17 years), which consists of children and
93 adolescents referred to a clinic for having problems with attention, learning and memory
94 (Holmes et al. 2018), and the Enhanced Nathan Kline Institute – Rockland Sample (NKI-RS, $N =$
95 335, aged 6 - 17 years), a community-ascertained sample (Nooner et al. 2012).

96 To investigate the neurocognitive architecture of g_f in a principled way, we used a watershed
97 model of individual differences. Based on the metaphor of a watershed, the model predicts a
98 hierarchical many-to-one mapping of partially independent effects such that upstream

99 tributaries (e.g. brain structure) contribute to intermediate cognitive phenomena (cognitive
100 endophenotypes, e.g. working memory and processing speed), which then contribute to
101 downstream, complex cognitive phenomena such as g_f (Cannon and Keller 2006; Kievit, Davis,
102 Griffiths, Correia, CamCAN, et al. 2016). See Figure 1 for a representation of the model.



103
104 *Figure 1. The Watershed Model.* Schematic representation of the watershed model
105 developed by (Cannon and Keller 2006) and adapted for the present study. Fluid ability is
106 hypothesized to be the downstream product of working memory and processing speed,
107 which are, in turn, the product of white matter contributions. Figure adapted from Kievit et
108 al. (2016).

109 SEM, as a statistical technique, is uniquely suited to modeling the kinds of complex
110 multivariate brain-behavior associations posited by the watershed model (Kievit et al. 2011;
111 Kline 2015). SEM combines factor analysis and path analysis (a variant of regression analysis).
112 It can model abstract cognitive constructs like g_f , by estimating latent variables from
113 observed task scores (i.e. manifest variables). This feature of SEM allowed us to model g_f ,
114 working memory, and processing speed in two independent samples, and thereby provided a
115 direct test of the generalizability of our findings. Second, SEM can test the simultaneous

116 relations between multiple cognitive and neural variables, allowing us to address the relative
117 contributions of different white matter tracts and different cognitive endophenotypes to g_f .
118 Finally, using SEM Trees (Brandmaier et al. 2013), a novel, decision-tree-based extension of
119 SEM, we investigated whether the associations in the watershed model change with age.

120 Based on the watershed model we made the following preregistered predictions
121 (<http://aspredicted.org/blind.php?x=u5pf6z>):

- 122 1. Working memory, g_f and processing speed are separable constructs.
- 123 2. Individual differences in g_f are predicted by working memory and processing speed.
- 124 3. White matter microstructure is a multi-dimensional construct.
- 125 4. There is a hierarchical relationship between white matter microstructure, cognitive
126 endophenotypes (working memory and processing speed) and g_f , such that white
127 matter contributes to working memory and processing speed, which, in turn
128 contribute to g_f .
- 129 5. The contribution of working memory and processing speed to g_f changes with age.

130 **Materials and Methods**

131 **Samples**

132 We analysed data from the CALM and NKI-RS sample, as described in detail by (Holmes et al.
133 2018) and (Nooner et al. 2012) respectively. See also Simpson-Kent et al. (2019). We had also
134 preregistered to analyse data from the ABCD cohort (Volkow et al. 2018). The latter cohort
135 contains only data for 9 - and 10 - year olds at present, however, which limits comparability
136 to CALM and NKI-RS, and makes it unsuitable for investigations of developmental differences.
137 We therefore opted to not analyse ABCD data here and instead recommend a replication of

138 the analyses presented here in ABCD once longitudinal data is available. The CALM sample
139 consists of children and adolescents referred by health and educational professionals as
140 having difficulties in attention, learning and/or memory. The NKI-RS is a community-
141 ascertained, lifespan sample, and representative of the general population of Rockland, New
142 York, and the United States as a whole, in terms of ethnicity, socioeconomic status etc. For
143 NKI-RS, we included data for participants under the age of 18 only to match the age range of
144 CALM and excluded data that were completed more than half a year after enrolment. The
145 latter criterion was implemented to ensure that age at assessment did not differ
146 substantively between cognitive measures. The final samples included 551 participants from
147 CALM (30.85% female, aged 5.17 - 17.92 years, $N_{\text{Neuroimaging}} = 165$) and 335 participants from
148 NKI-RS (43.48% female, aged 6.06 - 17.92 years, $N_{\text{Neuroimaging}} = 67$). See Table 1 for prevalence
149 of relevant disorders and learning difficulties in the samples.

150 *Table 1. Prevalence of Relevant Disorders and Learning Difficulties in the CALM and NKI-RS*
151 *cohorts*

Variable	Percentage CALM	Percentage NKI-RS
ADHD	31.94	17.01
Dyslexia	5.81	5.67
Autism	6.72	0.60
Mood disorder	0.54	0.90
Anxiety disorder	2.36	18.21
Medicated ¹	10.53	17.01
Speech/language problems	38.11	19.40

152 *Note.* ¹ unspecified medication for NKI-RS, ADHD-medication for CALM

153 **Cognitive Tasks**

154 We included cognitive tasks measuring the domains of g_f , working memory or processing
155 speed for CALM and NKI-RS. See Table 2 for the complete list of tasks used, and the

156 Supplementary Methods for task descriptions. Supplementary Figure 1 and 2 show raw
 157 scores on all tasks. The tasks modelled here were preregistered for CALM but not NKI-RS.

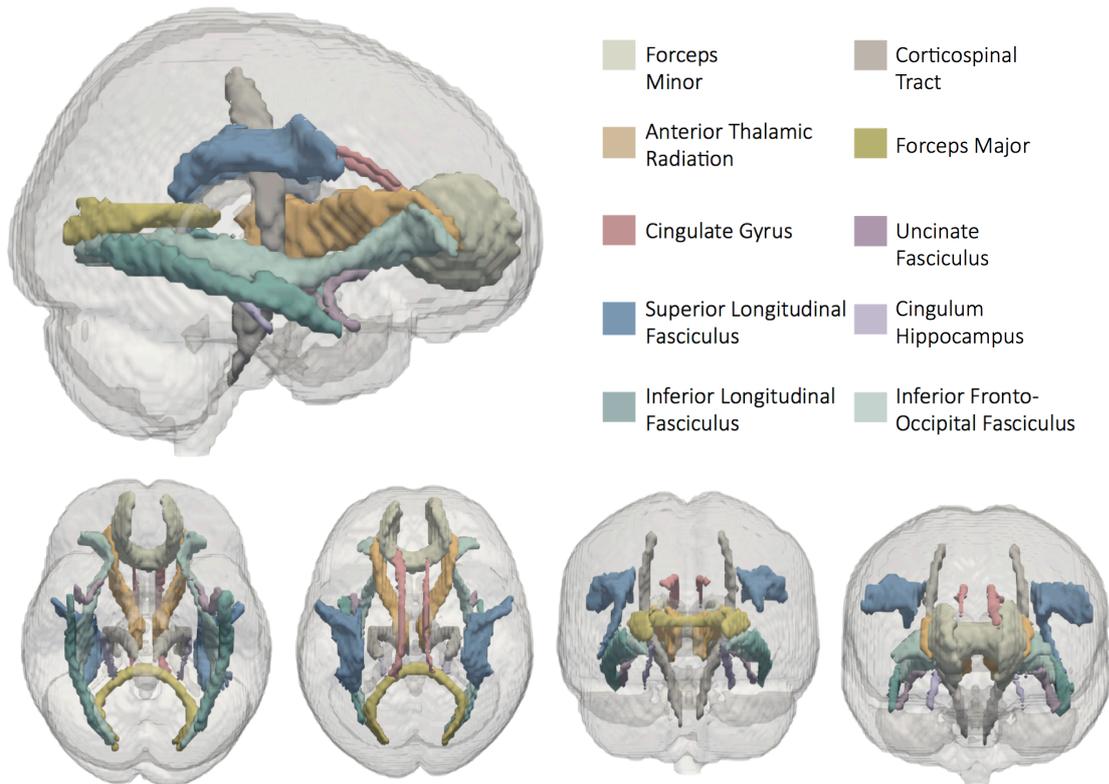
158 *Table 2. Cognitive Tasks Modelled*

	CALM	NKI-RS
Working memory	AWMA Digit Recall (forward digit span)	WISC-R Forward Digit Span
	AWMA Backward Digit Span	WISC-R Backward Digit Span
	AWMA Dot Matrix	-
	AWMA Mr X	-
	-	CNB N-back task
g_f	WASI-II Matrix Reasoning	WASI-II Matrix Reasoning
	-	WASI-II Block Design
	-	WASI-II Similarities
	-	CNB Verbal Reasoning
Processing speed	DKEFS Trail-Making	DKEFS Trail-Making
	PhAB Rapid Naming	-
	TEA-Ch RBBS	-
	-	CNB Motor Speed
	-	CNB Sensory Motor Speed

159 *Note.* See the Supplementary Methods for task descriptions. Abbreviations: AWMA -
 160 Automated Working Memory Assessment (Alloway 2007), CNB - Computerized
 161 Neurocognitive Battery (Gur et al. 2001), DKEF - Delis-Kaplan Executive Functioning System
 162 (Delis et al. 2004), PhAB – Phonological Assessment Battery (Gallagher and Frederickson
 163 1995), TEA-Ch RBBS - Test of Everyday Attention for Children, Red & Blues, Bags & Shoes
 164 subscale (Manly et al. 2001), WASI - Wechsler Abbreviated Scale of Intelligence - Second
 165 Edition (Wechsler 2011), WISC-R - Wechsler Intelligence Scale for Children – Revised
 166 (Kaufman 1975).

167 White Matter Microstructure

168 We modelled mean FA for all ten tracts of the Johns Hopkins University (JHU) white matter
 169 tractography atlas (Hua et al. 2008) averaged over the hemispheres (Figure 2). See
 170 Supplementary Methods for details of the MRI acquisition and processing and Supplementary
 171 Figure 3 and 4 for raw FA values in all tracts.



172
173 *Figure 2. White Matter Tracts Modelled in the Analyses.*

174 Analysis Methods and Structural Equation Modelling

175 Covariance matrices and scripts replicating key analyses can be obtained from:
176 https://github.com/df1234/gf_development. Supplementary Figure 5 and 6 show correlation
177 matrices of all tasks and white matter tracts modelled. We modelled raw scores for g_f and
178 working memory tasks, as preregistered. Raw scores on processing speed tasks were
179 transformed. This step was not preregistered, but found necessary to achieve model
180 convergence to ensure interpretability of scores. First, we inverted response time scores
181 (using the formula $y = 1/x$) to obtain more intuitive measures of ‘speed’ for all but the CNB
182 Motor Speed task, for which raw scores were already a measure of speed. Afterwards, we
183 applied a log-transformation to reaction time tasks to increase normality and aid estimation.
184 For the CNB Motor Speed task only, we additionally removed values $\pm 2 SD$ of the mean ($N =$
185 6) because the presence of these outliers had caused convergence problems.

186 We modelled the associations between cognition and white matter microstructure using SEM
187 in R (R core team 2015) using the package lavaan (Rosseel 2012). All models were fit using
188 maximum likelihood estimation with robust Huber-White standard errors and a scaled test
189 statistic. Missing data was addressed using full information maximum likelihood estimation.

190 We used SEM Trees to investigate whether the associations among cognitive and neural
191 measures differed with age. SEM Trees use decision tree methods to hierarchically split a
192 dataset into subgroups if parameter estimates differ significantly based on a covariate of
193 interest - in this case age (Brandmaier et al. 2013). We first ran the watershed model in
194 OpenMx (Boker et al. 2011) and then passed this model object to semtree to compute the
195 SEM Trees. We ran one SEM Tree for each parameter of interest (e.g. the covariance
196 between working memory and processing speed). All other parameters in each semtree
197 object were set to be invariant across groups to ensure that splits were specific to the
198 parameter of interest. We used a 10 - fold cross-validation estimation method as recommend
199 by (Brandmaier et al. 2013). For the path from the cingulate to working memory only we
200 used 5 - fold cross-validation because the model did not converge using 10 - fold cross-
201 validation. Minimum sample size in age group was set to $N = 50$ to ensure reliable estimation
202 of standard errors. Note that this choice effectively limited search space for potential splits to
203 ages 6.58 - 12.42 years for CALM and 8.08 - 15.49 years for NKI-RS.

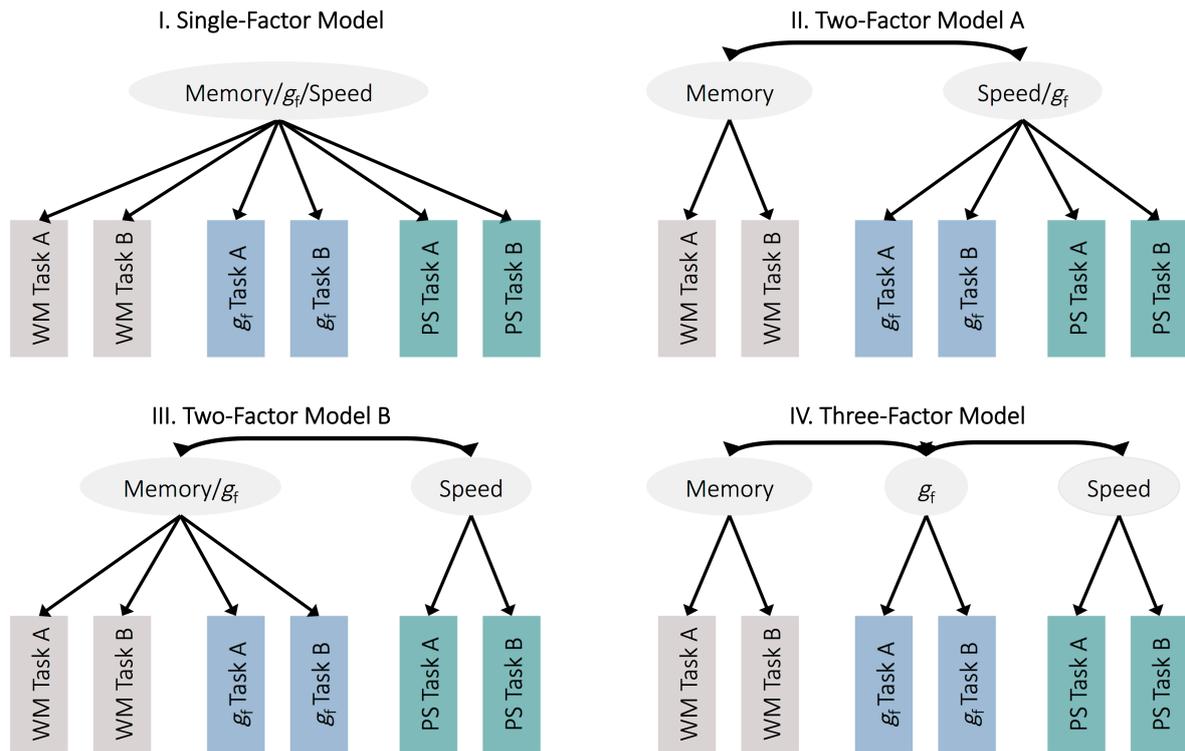
204 **Results**

205 To evaluate the hypotheses generated by the watershed model, we built up the watershed
206 model in steps and carried our comprehensive tests of model fit at each step. First, we
207 assessed the overall fit of our models to the data using the chi-square test, root mean square

208 error of approximation (RMSEA), comparative fit index (CFI) and standardized root mean
209 square residual (SRMR). Good absolute fit was defined as $RMSEA < 0.05$, $CFI > 0.97$ and $SRMR$
210 < 0.05 ; acceptable fit as $RMSEA = 0.08 - 0.05$, $CFI = 0.95 - 0.97$, $SRMR = 0.05 - 0.10$
211 (Schermelleh-Engel et al. 2003). Second, we assessed specific predictions from our models by
212 comparing them to alternative models. Comparative model fit for nested models was
213 assessed using the chi-square difference test. Non-nested models were compared using the
214 Akaike (AIC) weights, which indicates the probability of a model being the data-generating
215 model compared to all other models tested (Wagenmakers and Farrell 2004). Lastly, we
216 evaluated the significance and strength of relationships between specific variables in our
217 models by inspecting the Wald test for individual parameters, noting the joint R^2 where
218 relevant and reporting standardized parameter estimates. Absolute standardized parameter
219 estimates above 0.10 were defined as small effects, 0.20 as typical and 0.30 as large (Gignac
220 and Szodorai 2016).

221 **The Measurement Model of Cognition**

222 To examine the neurocognitive architecture of g_f , we started by modelling the cognitive
223 components of the watershed model: g_f , working memory and processing speed. Specifically,
224 we fit a three-factor model of cognition (Figure 3) and compared it to alternative
225 measurement models. This approach allowed us to test Hypothesis 1: namely that g_f , working
226 memory and processing speed form three separable, albeit likely correlated cognitive factors.
227 The Three-Factor Model (Figure 3) showed excellent *absolute fit* for both the CALM and NKI-
228 RS sample (Table 3), indicating that overall, the data was compatible with a model of g_f ,
229 working memory and processing speed as three separate factors.



230

231 *Figure 3. Different Measurement Models of Cognition.* Abbreviations: WM: working memory,
 232 PS: processing speed

233 The Three-Factor Model also showed very good *comparative fit* for NKI-RS, with a 96.60%
 234 probability of being the data-generating model compared to all alternative models tested, as
 235 indicated by its AIC weight (Figure 3). The evidence was more mixed for CALM, for which the
 236 Three-Factor Model showed a 27.15% probability of being the data-generating model, while
 237 Two-Factor Model B (Figure 3, treating working memory and g_f as a unitary factor) showed a
 238 72.85% probability of being the data-generating model, highlighting a close relationship
 239 between g_f and working memory for this sample. The Single-Factor Model and Two-Factor
 240 Model A (Figure 3, treating speed and g_f as a unitary factor) showed a very low
 241 (approximately 0%) probability of being the data-generating model, indicating that speed and
 242 g_f were clearly separable in both samples.

243 *Table 3. Model Fit of Competing Measurement Models*

Single-Factor Model	Two-Factor Model A	Two-Factor Model B	Three-Factor Model
<i>CALM</i>			
$\chi^2(20) = 70.28, p < .001$	$\chi^2(19) = 67.99, p < .001$	$\chi^2(19) = 41.66, p = .002$	$\chi^2(18) = 41.74, p = .001$
RMSEA = .068	RMSEA = .068	RMSEA = .047	RMSEA = .049
[.051-.085]	[.052 - .086]	[.027 - .066]	[.030 - .068]
CFI = .963	CFI = .964	CFI = .983	CFI = .983
SRMR = .047	SRMR = .043	SRMR = .032	SRMR = .032
AIC = 9697.18	AIC = 9696.44	AIC = 9668.58	AIC = 9670.55
BIC = 9800.66	BIC = 9804.24	BIC = 9776.37	BIC = 9782.66
AIC _{weight} = 0%	AIC _{weight} = 0%	AIC _{weight} = 72.85%	AIC _{weight} = 27.15%
<i>NKI-R</i>			
$\chi^2(35) = 109.96, p < .001$	$\chi^2(34) = 108.15, p < .001$	$\chi^2(34) = 64.85, p = .001$	$\chi^2(32) = 54.15, p = .009$
RMSEA = .080	RMSEA = .081	RMSEA = .052	RMSEA = .045
[.064 - .097]	[.064 - .098]	[.033 - .071]	[.024 - .065]
CFI = .936	CFI = .936	CFI = .974	CFI = .981
SRMR = .045	SRMR = .044	SRMR = .035	SRMR = .030
AIC = 7155.64	AIC = 7155.74	AIC = 7109.43	AIC = 7102.74
BIC = 7270.07	BIC = 7273.98	BIC = 7227.67	BIC = 7228.60
AIC _{weight} = 0%	AIC _{weight} = 0%	AIC _{weight} = 3.40%	AIC _{weight} = 96.60%

244 *Note.* See Figure 3 for the configuration of different models. Abbreviations: Akaike Information Criterion (AIC),
 245 Bayesian Information Criterion (BIC), Akaike weight (AIC_{weight})

246 Overall, these result provide mixed evidence for Hypothesis 1: Even though working memory,
 247 processing speed and g_f were highly correlated in both samples (Table 4), processing speed
 248 formed a clearly separable factor from working memory and g_f in both samples. Working and
 249 g_f , however, were clearly separable only in NKI-RS, but not CALM, suggesting greater
 250 similarity between g_f and working memory in the CALM sample. To facilitate comparison
 251 across samples and in accordance with our preregistered analysis plan we nonetheless used
 252 the three-factor measurement model (Table 4, Supplementary Table 1) in all subsequent
 253 analyses.

254 *Table 4. Covariance between Cognitive Measures in the Three-Factor Model*

Sample	Path	Standardized Estimate
CALM	$g_f \leftrightarrow$ memory	0.71, $z = 28.42$, $p < .001$
	$g_f \leftrightarrow$ speed	0.55, $z = 12.20$, $p < .001$
	memory \leftrightarrow speed	0.79, $z = 19.35$, $p < .001$
NKI-RS	$g_f \leftrightarrow$ memory	0.91, $z = 19.51$, $p < .001$
	$g_f \leftrightarrow$ speed	0.81, $z = 24.73$, $p < .001$
	memory \leftrightarrow speed	0.87, $z = 17.43$, $p < .001$

255 *Note.* See Supplementary Table 1 for factor loadings.

256 **The Relationship between Working Memory, Processing Speed and g_f**

257 We next examined the relationships between working memory, processing speed and g_f in
 258 more detail. Specifically, we fit a SEM including regression paths between working memory
 259 and g_f , as well as speed and g_f , to test Hypothesis 2, that working memory and processing
 260 speed each predict individual differences in g_f . We found that this model showed good
 261 absolute fit for both samples (CALM: $\chi^2(18) = 41.74$, $p = .001$; RMSEA = .049 [.030 - .068]; CFI
 262 = .983; SRMR = .032, NKI-RS: $\chi^2(32) = 54.15$, $p = .009$; RMSEA = .045 [.024 - .065]; CFI = .981;
 263 SRMR = .030), indicating that, overall, the data was compatible with our model.

264 To further scrutinize the relationship between g_f , working memory and speed, we compared
 265 our freely-estimated model to a set of alternative models with different constraints imposed
 266 upon the regression paths. First, to test whether working memory and speed each made
 267 different contributions, we tested an alternative model in which the paths from processing
 268 speed and working memory to g_f were constrained to be equal. In CALM ($\Delta\chi^2(1) = 15.53$, $p <$
 269 $.001$), but not NKI-RS ($\Delta\chi^2(1) = 3.25$, $p = .072$), the freely-estimated model fit better than the
 270 equality-constrained model, indicating that working memory and speed each made different
 271 contributions in CALM but not NKI-RS. Next, we tested whether the freely estimated model
 272 fit better than a model in which the path between g_f and working memory was constrained

273 to zero. We found that that the freely estimated model fit better for both samples (CALM:
274 $\Delta\chi^2(1) = 20.77, p < .001$; NKI-RS: $\Delta\chi^2(1) = 12.97, p < .001$). In line with our hypothesis, this
275 result indicates that working memory makes a significant incremental contribution to g_f .
276 Finally, we tested a model in which the path between g_f and processing speed was
277 constrained to zero. This model showed no difference in fit to the freely estimated model for
278 CALM ($\Delta\chi^2(1) = 0.02, p = .875$) or NKI-RS ($\Delta\chi^2(1) = 0.04, p = .849$). Contrary to our hypothesis,
279 this indicates that there was no clear incremental contribution of processing speed to g_f .

280 Finally, we inspected standardized path estimates of the freely estimated model to assess the
281 effect sizes of working memory and processing speed. Parameter estimates showed that
282 working memory showed a greater effect on g_f than processing speed, particularly in CALM
283 (Table 5) even though raw correlations between g_f and speed were high in both samples
284 (Table 4).

285 *Table 5. Regression Path Estimates.*

Sample	Path	Standardized Estimate
CALM	speed -> g_f	-0.01, $z = -0.16, p = .876$
	memory -> g_f	0.72, $z = 7.65, p < .001$
NKI-RS	speed -> g_f	0.06, $z = 0.21, p = .208$
	memory -> g_f	0.86, $z = 1.81, p = .070$

286
287 Overall these results provide mixed evidence for Hypothesis 2: There was good evidence that
288 working memory and speed made a significant joint contribution to g_f , and that working
289 memory made an incremental contribution to g_f in CALM. Contrary to our hypothesis, and
290 the watershed model, however, processing speed showed no significant incremental
291 contribution to g_f , above and beyond working memory. We explore likely explanations for this
292 finding in the Discussion.

293 The Measurement Model of White Matter

294 We next examined the measurement model of white matter to test Hypothesis 3, namely
295 that white matter microstructure is a multi-dimensional construct. Specifically, we tested
296 whether white matter microstructure could be adequately captured by a single factor by
297 examining absolute model fit. As expected, the single-factor model of white matter
298 microstructure did not fit the data well (CALM: $\chi^2(35) = 124.63$, $p < .001$; RMSEA = .125 [.103
299 - .147]; CFI = .933; SRMR = .039; NKI-RS: $\chi^2(35) = 132.33$, $p < .001$; RMSEA = .204 [.167 - .242];
300 CFI = .885; SRMR = .023). This indicates that white matter microstructure could not be
301 reduced to a single 'global FA' dimension in our samples, in line with (Lövdén et al. 2013;
302 Kievit, Davis, Griffiths, Correia, CamCAN, et al. 2016) and supporting Hypothesis 3. We
303 therefore modelled each of the ten white matter tracts separately in all subsequent models.

304 The Watershed Model: Relationships between Cognition and White Matter

305 Next, we fit the full watershed model including white matter, working memory, processing
306 speed and g_f . Following our general analysis procedure, we investigated overall model fit,
307 alternative models and individual path estimates to gain a comprehensive understanding of
308 the relationships in the watershed model and to test Hypothesis 4 - that white matter
309 contributes to working memory capacity and processing speed, which, in turn, contribute to
310 g_f .

311 We found largely converging results across samples. The watershed model showed good
312 absolute fit in CALM ($\chi^2(78) = 107.78$, $p = .014$; RMSEA = .026 [.012 - .038]; CFI = .981; SRMR
313 = .043) and acceptable fit in NKI-RS ($\chi^2(112) = 219.22$, $p < .001$; RMSEA = .053 [.043 - .064];
314 CFI = .928; SRMR = .088). White matter explained large amounts of variance in working
315 memory ($R^2_{\text{CALM}} = 32.3\%$; $R^2_{\text{NKI-RS}} = 46.1\%$) and processing speed ($R^2_{\text{CALM}} = 38.2\%$; $R^2_{\text{NKI-RS}} =$

316 54.4%), which, in turn, explained even more variance in g_f ($R^2_{\text{CALM}} = 51.2\%$; $R^2_{\text{NKI-RS}} = 78.3\%$).

317 In line with Hypothesis 4, this indicates that the watershed model fit the data overall.

318 Comparing the freely estimated watershed model to alternative, constrained, models

319 showed that white matter contributed significantly to memory and processing speed.

320 Specifically, a model in which paths from white matter to processing speed were constrained

321 to zero fit worse than the freely-estimated model (CALM: $\Delta\chi^2(10) = 50.26$, $p < .001$; NKI-RS:

322 $\Delta\chi^2(10) = 27.19$, $p = .002$), as did a model in which paths from white matter to working

323 memory were constrained to zero (CALM: $\Delta\chi^2(10) = 52.26$, $p < .001$; NKI-RS: $\Delta\chi^2(10) = 25.85$,

324 $p = .004$). As hypothesised, white matter therefore contributed to both processing speed and

325 working memory.

326 We next inspected that relationship between individual white matter tracts and working

327 memory and speed in more detail. A model in which paths from white matter to working

328 memory and speed were constrained to be equal, fit worse than the freely-estimated

329 watershed model for CALM ($\Delta\chi^2(18) = 47.76$, $p < .001$) and NKI-RS ($\Delta\chi^2(18) = 30.42$, $p = .034$),

330 indicating that the role of white matter microstructure in supporting working memory and

331 processing speed differed across tracts. This supports the notion that there is a many-to-one

332 mapping between white matter and cognition - a core tenet of the watershed model.

333 Investigating individual standardised parameter estimates of the different white matter tracts

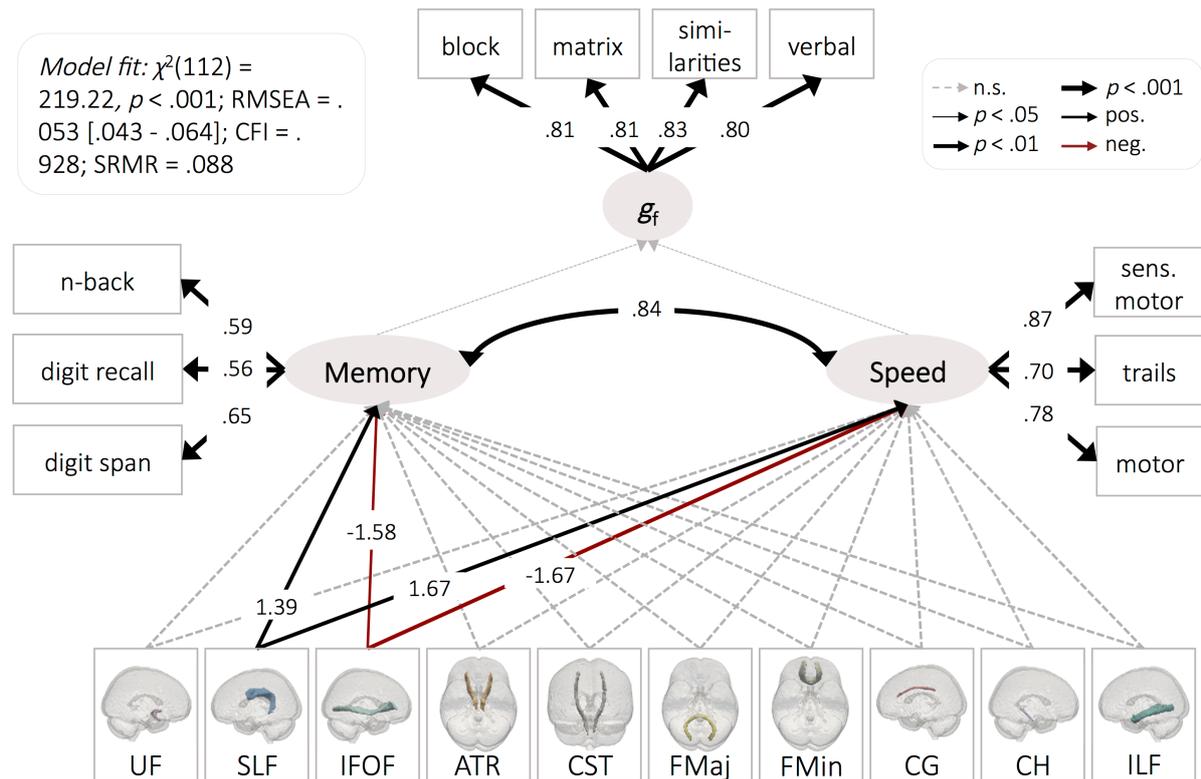
334 showed that for CALM, only the anterior thalamic radiation contributed significantly to

335 processing speed, whereas the superior longitudinal fasciculus, forceps major and cingulum

336 were significantly, independently and positively related to working memory (Figure 4). For

337 NKI-RS, the superior longitudinal fasciculus was significantly and positively related to

338 processing speed and working memory (Figure 5). Two tracts showed an unexpected,



356

357 *Figure 5. The Watershed Model in NKI-RS. See Supplementary Table 3 for regression*
 358 *estimates. Residual covariances between white matter tracts were allowed but are not*
 359 *shown for simplicity.*

360 Finally, we probed the watershed model in more detail by testing a set of alternative

361 expressions of the watershed model still compatible with the core tenants of the watershed

362 model – as well as a set of alternative models incompatible with the watershed model. We

363 compared all alternatives (see Figure 6 for graphical representations) to the original

364 watershed model by inspecting their relative probability of being the data-generating model

365 as indicated by their AIC weights (Wagenmakers and Farrell 2004). We found that the original

366 watershed model showed a very high probability (98.58%) of being the data-generating

367 model for CALM but only a 0.10% probability for NKI-RS. For NKI-RS, a different expression of

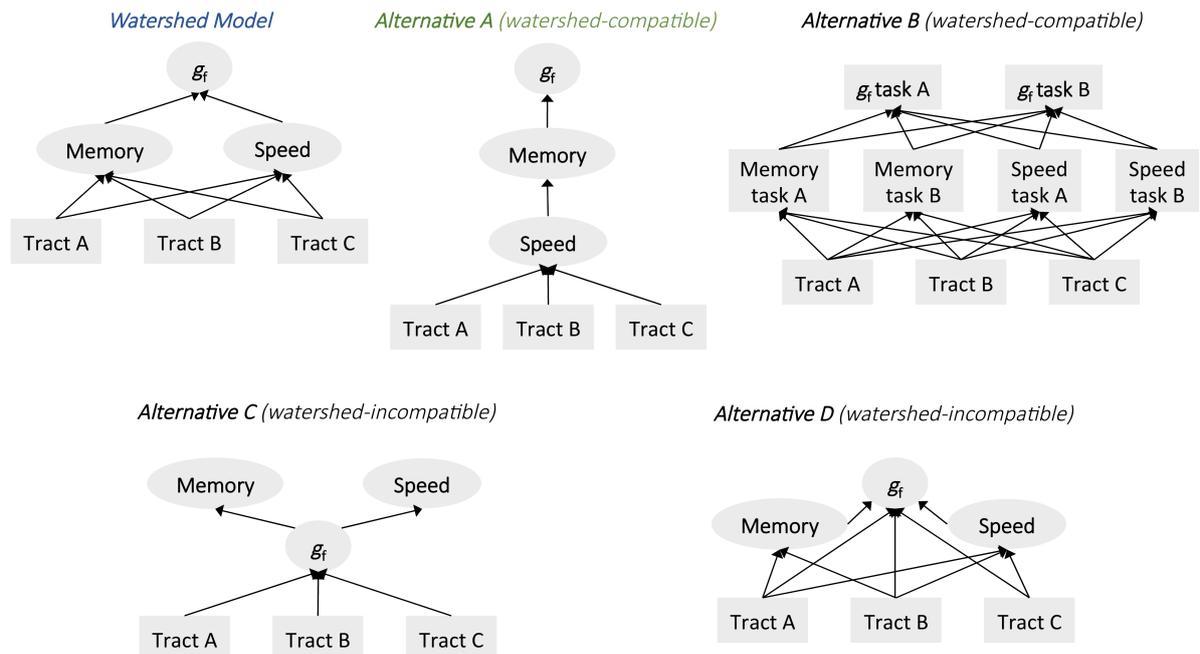
368 the watershed model, such that g_f was regressed on working memory, which was regressed

369 on processing speed, which was then regressed on white matter (Alternative A, Figure 6)

370 showed a 95.04% probability of being the data-generating model. This model only showed a

371 0.37% probability for CALM. Another expression of the watershed model, in which all tasks
372 were modelled separately as manifest, rather than latent, variables (Alternative B, Figure 6),
373 showed no advantage over the watershed model for CALM (0.00% probability) or NKI-RS
374 (0.00% probability). We next tested two alternative models incompatible with the tenants of
375 the watershed model. We found that a model in which the hierarchy between cognitive
376 endophenotypes and g_f was inverted (Alternative C, Figure 6) showed comparatively low
377 probability of being the data-generating model for both CALM (0.00%) and NKI-RS (2.86%).
378 Similarly, a model in which g_f was directly regressed on white matter, working memory and
379 processing speed (Alternative D, Figure 6), showed no clear advantage over the watershed
380 model for CALM (1.05% probability) or NKI-RS (0.00% probability). Overall these model
381 comparisons highlight that while the watershed model fit the data for both samples and had
382 large explanatory power (as indicated by R^2 s), the precise configuration of the watershed
383 model may differ somewhat between cohorts.

384 In summary, we found that the watershed model performed well overall for both cohorts. As
385 hypothesised, white matter contributed to working memory and processing speed, which, in
386 turn, contributed to g_f , and explained large amounts of variance therein. Also as predicted by
387 the watershed model, there was a many-to-one mapping between white matter tracts and
388 cognition. The exact configuration of the watershed model, however, may differ slightly
389 between cohorts. These differences may be a function of cohort differences in sample size,
390 average levels of cognitive ability and/or the specific tasks used – a topic we will return to in
391 the Discussion.



392

393 *Figure 6. Configuration of Alternative Models.* Alternatives A and B are watershed-
 394 compatible, while C and D are watershed-incompatible. The best-fitting model for CALM is
 395 highlighted in blue; the best-fitting model for NKI-RS is highlighted in green. Regression paths
 396 only are shown for simplicity. Square shapes denote manifest variables and oval shapes
 397 latent variables.

398

399 Testing for potential confounds

400 We carried out a series of supplementary and non-preregistered analyses to examine
 401 whether possible confounders influenced our models. These analyses showed that our
 402 findings were robust to the inclusion of covariates such as scanner motion or socio-economic
 403 status. They were also robust across genders and participants taking or not taking
 404 medication. There were no differences in the structure of the model between participants
 405 with and without diagnosed disorders for CALM. Potential small differences cannot be ruled
 406 out for NKI-RS, likely due to the low number of diagnosed participants of $N = 106$
 407 (Supplementary Material).

408 Age-Related Differences in the Neurocognitive Architecture of g_f

409 Finally, we tested Hypothesis 5 - that that the contribution of working memory and
410 processing speed to g_f varied with age. We first inspected cross-sectional differences in g_f ,
411 working memory and processing speed, and then used SEM trees to investigate potential
412 age-differences in the relationships between these factors. In additional, non-preregistered,
413 analyses we also used SEM Trees to investigate potential age-differences in the relationship
414 between white matter and cognitive endophenotypes by inspecting paths that were
415 significant in the watershed model (Figure 4 and 5).

416 SEM trees combine SEMs with decision tree methods, separating a dataset into subgroups (in
417 this case age groups) if SEM parameter estimates of interest differ sufficiently (Brandmaier et
418 al. 2013). SEM trees allowed us to investigate age as a potential moderator without imposing
419 a-priori categorical age splits. We initially allowed for no more than two age groups. This
420 yielded inconsistent results for CALM and NKI-RS (Supplementary Table 4). To test whether
421 these inconsistencies were an artefact of allowing for only two groups, we repeated our
422 analysis and allowed for up to four age groups. This analysis yielded consistent results
423 between CALM and NKI-RS (Table 6). This pattern of results indicates that the initial
424 parameters of our analysis caused us to miss relevant age differences.

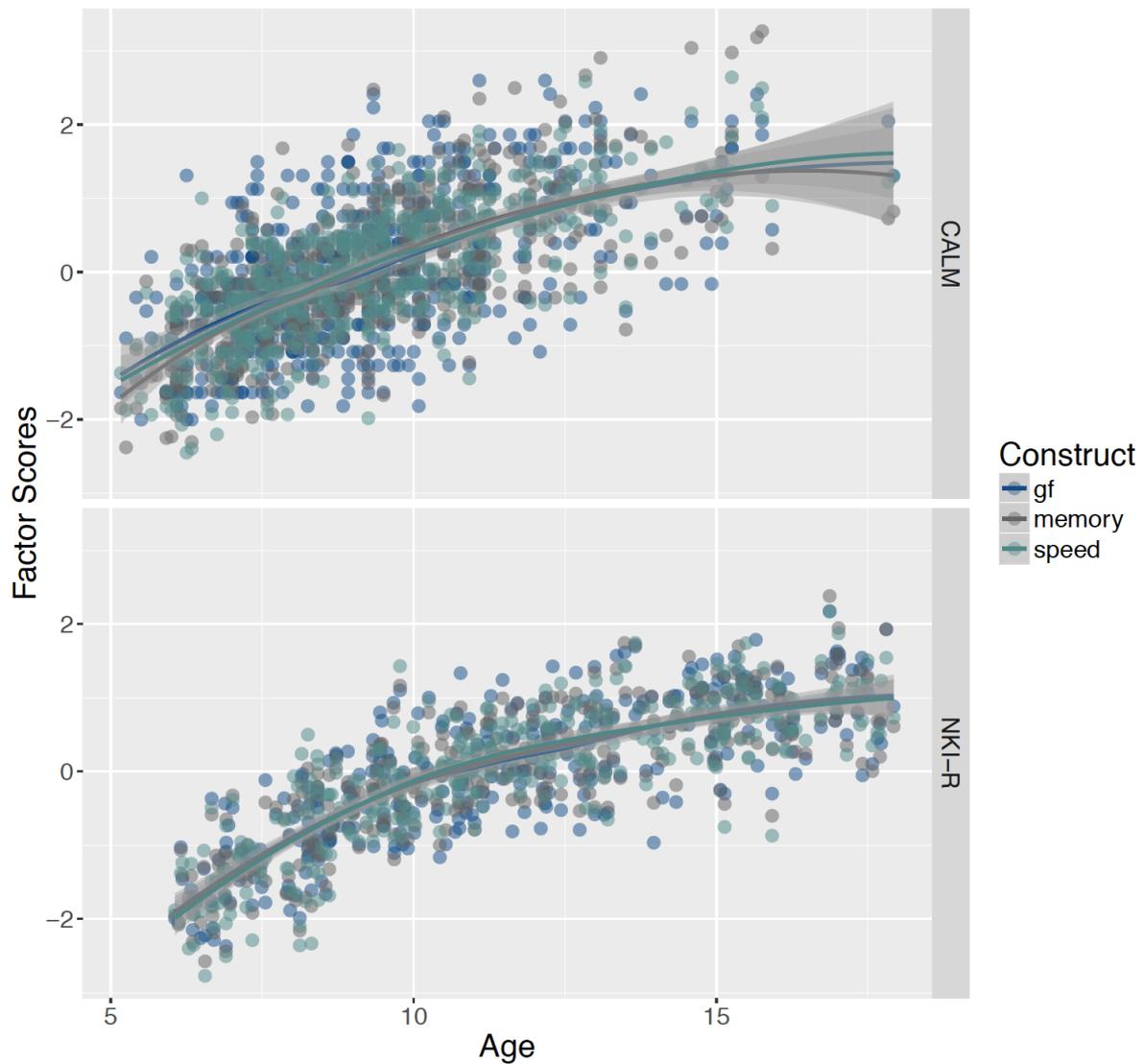
425 As shown in Figure 7, g_f , working memory and processing speed factor scores increased with
426 age for all three cognitive phenotypes. In line with our hypothesis, SEM trees showed that
427 there were pronounced age-related differences in brain-behaviour in childhood and
428 adolescence (Table 6). For both samples and all but one path, there was an initially strong
429 relationship between components of the watershed model, then a dip around ages 7 - 9
430 years for CALM and age 8 for NKI-RS, followed by an increase in path strength around ages 11

431 – 12 (see Supplementary Figure 7 for a graphical representation of these results).
 432 Speculatively, this pattern of results is consistent with an interpretation of a reorganization of
 433 neurocognitive faculties in late childhood, followed by a consolidation of neurocognitive
 434 pathways around the onset of adolescence (Johnson 2000, 2011).

435 *Table 6. SEM Tree Results for the Watershed Model.*

Path	Est. Before	Age Split 1	Est. Betw.	Age Split 2	Est. Betw.	Age Split 3	Est. After
<i>CALM</i>							
memory <--> speed	0.85	8.46	0.97	9.46	0.74	-	-
memory → g_f	0.83	9.38	0.42	10.04	1.14	10.88	0.94
speed → g_f	0.04	6.88	-0.19	11.21	0.17	-	-
SLF → memory	0.67	7.21	0.18	11.21	0.76	-	-
FMaj → memory	0.59	7.71	0.14	9.29	0.33	11.13	0.74
CG → memory ¹	0.64	6.96	0.09	11.04	0.70	-	-
ATR → speed	0.96	7.13	0.68	7.96	0.17	11.96	0.65
<i>NKI-RS</i>							
memory <--> speed	0.90	9.82	0.48	14.72	1.11	-	-
memory → g_f	1.10	8.59	0.59	12.67	1.03	-	-
speed → g_f	0.53	8.59	-0.12	12.96	0.52	-	-
SLF → memory	2.15	8.30	1.47	12.15	1.93	-	-
SLF → speed	3.12	8.63	1.83	15.09	2.31	-	-

436 *Note.* The table shows differences in parameter estimates for paths of interest (as shown in
 437 Figure 4 and 5) depending on participants' age in years. Our analyses allowed for a maximum
 438 of three age splits (and thus four age groups). An absence of a third age split (denoted by '-'
 439 in the table), indicates that the SEM tree split only twice, suggesting no further changes in
 440 parameter strength after the second split. See Supplementary Figure 7 for a graphical
 441 representation of these results.



442

443 *Figure 7. Cognitive Factor Scores by Age.*

444

Discussion

445 We here used multivariate statistical techniques to investigate the neurocognitive

446 architecture of g_f in two large ($N_{\text{CALM}} = 551$, $N_{\text{NKI-RS}} = 335$) developmental cohorts and, for the

447 first time, investigated how the neurocognitive architecture of g_f changes dynamically with

448 age. We tested a preregistered watershed model of g_f , which predicts a hierarchy of partially

449 independent effects. As might be expected from a multi-cohort study, there were some

450 differences between the community-ascertained cohort (NKI-RS) and the cohort of children

451 and adolescents with learning difficulties (CALM) in specific path estimates. Overall however,
452 we found strikingly convergent results across these two heterogeneous samples. The
453 watershed model performed well for both CALM and NKI-RS: White matter contributed to
454 working memory and processing speed, which, in turn, contributed to g_f and explained 51%
455 of variance therein for the CALM sample and 78% of variance for NKI-RS. Models were robust
456 across genders, participants taking or not taking medication and when controlling for socio-
457 economic status and scanner motion. Investigations of age effects showed that the
458 relationship between cognitive abilities and white matter dipped in strength around ages 7-
459 12 years. Speculatively, this age-effect may reflect a reorganization of the neurocognitive
460 architecture during pre-puberty and early puberty (Byrne et al. 2017). These findings have
461 implications for understanding and targeting cognitive impairments in populations with
462 learning difficulties.

463 The watershed model tested here consists of three levels: g_f forms the most down-stream
464 point, with working memory and processing speed as intermediate tributaries, and white
465 matter microstructural tracts as upstream sources. Previous studies suggested that matter
466 microstructure is best characterised by a single, 'global FA' factor (Penke et al. 2010) while
467 others have contended that association patterns among different white matter tracts are
468 more complex (Lövdén et al. 2013; Kievit, Davis, Griffiths, Correia, CamCAN, et al. 2016). Here
469 we found strong evidence for a multifactorial view of white matter tracts – for both samples,
470 a unidimensional model of white matter fit poorly, and for CALM, multiple tracts also showed
471 partially independent contributions to distal cognitive outcomes. This is in line with the
472 watershed model. There were some differences between cohorts as to which tracts
473 contributed most to working memory and processing speed: In line with previous research

474 (Kievit, Davis, Griffiths, Correia, CamCAN, et al. 2016; MacPherson et al. 2017; Bathelt et al.
475 2018), we found that the anterior thalamic radiation was related to processing speed, as
476 were the forceps major, forceps minor and the cingulum to working memory for CALM.
477 However, these tracts were not significant for NKI-RS. These differences between samples
478 may reflect differences in brain-behaviour mapping between more atypical and typical
479 cohorts (Bathelt et al. 2018), or simply sampling variance across two independent cohorts
480 where one cohort (NKI-RS) has a more modest number of participants with imaging data. Of
481 note, however, the superior longitudinal fasciculus was consistently associated with working
482 memory in both samples. For NKI-RS, the superior longitudinal fasciculus was also associated
483 with processing speed. The superior longitudinal fasciculus is a large, bilateral association
484 fibre connecting temporal, occipital, parietal and frontal regions (Kamali et al. 2014). It is
485 therefore well-situated for supporting cognitive processes such as g_f , which rely on
486 integrative multiple-demand systems (Jung and Haier 2007; Fedorenko et al. 2013; Parlatini
487 et al. 2017).

488 The cognitive levels of the watershed model highlighted a close relationship between
489 working memory and g_f . Previous studies had variably suggested that g_f and working memory
490 (Kyllonen and Christal 1990; Fukuda et al. 2010), or g_f and processing speed (Kail and
491 Salthouse 1994; Salthouse 1996; Coyle et al. 2011; Ferrer et al. 2013) may be most closely
492 related. We found that all three cognitive factors were highly correlated for both samples.
493 Nonetheless, processing speed formed a cognitive factor clearly separable from working
494 memory and g_f . Working memory and g_f , in turn, were separable in the community-
495 ascertained NKI-RS but not in CALM, the cohort of children and adolescents with learning
496 difficulties. This close relationship between g_f and working memory was also evident in other

497 models of CALM where processing speed and working memory were used as joint predictors
498 of g_f . Contrary to our hypotheses, processing speed became non-significant after controlling
499 for working memory here. There are several possible, and not mutually exclusive,
500 explanations for this finding. First, and in line with previous work showing that time-
501 constraints increase isomorphism of g_f and working memory (Chuderski 2013), even standard
502 implementations of g_f tasks may place considerable time-pressure on struggling learners,
503 whereby increasing g_f - working memory covariance. Second, a broader set of speed tasks
504 (which might be captured by several latent variables for clerical speed tasks, choice reaction
505 time tasks and variability indices) might show higher predictive power than the single latent
506 variable for speed, which could be modelled here. Third, the watershed model might be
507 configured somewhat differently for some populations, such that speed forms an
508 intermittent level in the hierarchy between white matter and working memory (Alternative A,
509 Figure 6). There was some evidence for this in NKI-RS, indicating that the hierarchy of the
510 watershed model might be differentiated more in cohorts of older ages and/or higher ability
511 levels. We note that all of these explanations would still be compatible with the notions of
512 the watershed model, and remain to be teased apart by future research. For now, we suggest
513 that our findings support the notion that mental information processing capacity, as
514 measured by working memory, is a key determinant of g_f (Kyllonen and Christal 1990; Fukuda
515 et al. 2010).

516 The associations in the watershed model differed between ages in a complex, non-monotonic
517 fashion. Previous research had suggested either a decrease in covariance among cognitive
518 domains with age (age differentiation) (Garrett 1946), an increase in covariance with age (age
519 de-differentiation) (Blum and Holling 2017), or no changes with age (Tucker-Drob 2009; de

520 Mooij et al. 2018). These investigations have traditionally focussed on relations between
521 cognitive domains, however, not on relationships between brain and cognition - although
522 see de Mooij et al. (2018). Possible linear and non-linear changes in brain-behaviour mapping
523 with age have remained mostly unexplored (Tamnes et al. 2017). Using structural equation
524 modelling trees, a novel decision-tree-based technique, we here found evidence of complex
525 developmental differences that were consistent across samples and relationships in the
526 watershed model: Initially strong path estimates showed a pronounced decrease in strength
527 around ages 7 - 9 years, followed by a renewed increase in the strength, even surpassing
528 initial levels, around ages 10 - 15.

529 There are at least two possible explanations for this developmental dip in brain-cognition
530 relationships. First, there may be a true decrease in relationship strength during this time of
531 life. Possibly, other cognitive skills such as verbal reasoning temporarily support g_f , resulting
532 in weaker relationships between g_f and working memory, for instance. Alternatively, the
533 configuration of the watershed model may change temporarily during this time, which could
534 also manifest in an apparently weaker covariance structure. In this case, the true relationship
535 between g_f , memory, speed and white matter may still be strong, just configured differently
536 from the watershed model. Both explanations are compatible with the interactive
537 specialization theory (Johnson 2000, 2011), which predicts as remapping of the relationships
538 between brain substrates and cognitive abilities during development.

539 On a physiological level, this age effect may be driven by neuroendocrine changes during pre-
540 and early puberty. Puberty is driven by a complex and only partially understood set of
541 hormonal events including gonadarche and adrenarche (Sisk and Zehr 2005). Gonadarche
542 begins with the secretion of gonadotropin-releasing hormone from the hypothalamus around

543 ages 10-11 years and closely tracks the overt bodily changes of puberty (Dorn 2006).
544 Andrenarche, beginning with the maturation of the andrenal gland, starts as early as six years
545 of age, and is increasingly recognized as a complimentary driver of puberty and brain
546 development (Byrne et al. 2017). It is possible that the hormonal changes of andrenarche and
547 early gonadarche may lead to a level of neural reorganization, which may initially appear as
548 weaker relationships in the watershed model. The sweeping bodily, social and cognitive
549 changes happening in early adolescence may then drive a consolidation of the neurocognitive
550 architecture of g_f .

551 On a more general level, this age effect suggests the existence of potential non-linear
552 changes in brain-behaviour mapping during childhood and adolescence and underlines the
553 value of modern statistical approaches, such as SEM Trees, for the study of age-related
554 differences. Nonetheless, it is worth noting that these findings, based on an inherently
555 exploratory technique, will need to be replicated in future confirmatory studies with fine-
556 grained data on puberty and larger sample sizes. The latter will also allow for detailed
557 investigations of potential gender differences.

558 Testing our model in two different samples allowed us to address several critical questions:
559 First, participants from both samples completed a small set of common and a larger set of
560 different cognitive tasks. Therefore, the results obtained here are likely not task-specific, but
561 rather can be expected to generalize to the domains of working memory, processing speed
562 and g_f (Noack et al. 2014). For instance, we show that even though one cohort (NKI-RS)
563 performed mainly clerical speed tasks and the other (CALM) mainly choice reaction time
564 tasks, our finding that processing speed did not significantly contribute to g_f after controlling
565 for working memory, was robust across these samples and tasks. Second, by comparing a

566 community-ascertained sample (NKI-RS) and a sample of children and adolescents with
567 learning difficulties (CALM), we demonstrated that the watershed model performed well for
568 very different populations. We are not able to make general claims about potential
569 differences between more typical and atypical populations, however. CALM and NKI-RS were
570 collected in countries with somewhat different socio-economic conditions (United Kingdom
571 and United States of America), the samples were of different sample sizes, and participants'
572 ages were distributed more evenly in one cohort (NKI-RS) than the other (CALM). While DTI
573 images were processed with the same pipeline across sites, the scanner and MRI acquisition
574 protocol were different. Although previous work suggests that FA is relatively robust measure
575 in multi-site comparisons (Vollmar et al. 2010), we cannot rule out site differences as a
576 potential confound. It will therefore be necessary to replicate these findings in large typical
577 and atypical cohorts collected in the same setting.

578 Our study illustrates some of the advantages and challenges of preregistered secondary data
579 analyses. We agree with others in the field that secondary data analysis need not be and
580 should not be confounded with purely exploratory research (Mills and Tamnes 2014; Orben
581 and Przybylski 2019; Scott and Kline 2019). Preregistrations, as well as dedicated multivariate
582 methods such as SEM, can help to reduce the scope for analytic flexibility and increase
583 scientific rigour when using secondary datasets, which are often rich and multivariate in
584 nature. Preregistrations also do not preclude the use of exploratory methods (such as SEM
585 trees used here) or the ability to ask exploratory questions (such as looking at age differences
586 in the relationships between brain and cognition reported here) - they merely facilitate the
587 distinction between exploratory and confirmatory research (Wagenmakers et al. 2012). There
588 are, however, some unique challenges to preregistering secondary data analyses worth

589 noting. First, information on the precise measures collected is not always easily available
590 prior to data access, which can limit the level of detail in which an analysis can be
591 preregistered. Second, data quality and the level of data-processing (the latter is particularly
592 relevant for MRI data) is not always clear a priori (e.g. see Kievit et al. 2018), which can
593 necessitate changes to analyses plans after data inspection. Third, convergence issues are
594 fairly common when using complex multivariate methods such as SEM. We found it
595 necessary to transform some of our speed variables, for instance, to achieve model
596 convergence. Such post-hoc modifications, not guided by the palatability of the results, but
597 rather by unforeseen, and sometimes unforeseeable, practical considerations, mean that
598 preregistration can sometimes fall short of full compliance. Nevertheless, we believe that
599 even imperfect preregistrations, alongside shared code, data and the transparent
600 presentation of results, can help the reader distinguish between confirmatory and
601 exploratory results, and adjust their level of confidence in conclusions accordingly. For
602 guidance on maximizing transparency in preregistration of secondary data, see Weston et al.
603 (2018).

604 A key limitation of our study is that our samples were cross-sectional, and not longitudinal.
605 Although the relatively narrow age range makes large cohort effects unlikely, it may still be
606 that there were differences in recruitment and selection that varied across the age range.
607 Moreover, while we were able to investigate *individual differences* in g_i , we could not assess
608 *intra-individual changes* during childhood and adolescence. As such, the cross-sectional
609 nature of our samples limits our ability to make inferences about developmental dynamics.
610 Longitudinal data will also be necessary to scrutinize the causal flow of effects in the
611 watershed model.

612 The findings from our study have implications understanding and targeting cognitive
613 impairments in populations with learning difficulties. First, the close relationship between
614 working memory and g_f found here and in other studies (Fukuda et al. 2010; Chuderski 2013),
615 indicates that children and adolescents struggling with working memory are likely to also
616 struggle in terms of complex reasoning tasks. Either reducing working memory load,
617 decreasing time constraints, or training working memory and fluid ability capacity in such
618 populations may therefore be promising lines of inquiry for intervention studies. It is worth
619 highlighting, however, that cognitive training studies have so far shown little evidence of (far)
620 transfer: Training abstract reasoning, a common measure of g_f , has not resulted in robust
621 increases in working memory (Knoll et al. 2016) and working memory training has not been
622 shown to transfer to reasoning skills or school performance (Dunning et al. 2013;
623 Schwaighofer et al. 2015). Similarly, transfer from processing speed to reasoning seems to be
624 limited (Mackey et al. 2011). The results obtained here suggest that interventions may
625 increase their chance of success by implementing programs of sufficient complexity to affect
626 the entire neurocognitive architecture of effects (see also Kievit et al. 2016). The level of
627 intensity required to produce sustained benefits may need to be as demanding and
628 consistent as education itself, which shows robust effects in increasing general cognitive
629 abilities over time (Ritchie and Tucker-Drob 2018). This work and work by others (Noack et al.
630 2014) also highlights the value of assessing, modeling, and potentially intervening on,
631 multiple tasks, rather than relying on a single task to capture complex cognitive domains such
632 as g_f . Finally, the age-related differences in the relationships of the watershed model
633 observed using SEM-trees suggest that some interventions may work best at particular
634 developmental phases.

635

Acknowledgements

636 D.F. and J.B. are supported by the UK Medical Research Council (MRC). I.L.S.-K. is supported
637 by the Cambridge Trust. R.A.K. is supported by the Sir Henry Wellcome Trust (107392/Z/15/Z)
638 and MRC Programme Grant (MC-A060-5PR60). The Centre for Attention Learning and
639 Memory (CALM) research clinic is based at and supported by funding from the MRC
640 Cognition and Brain Sciences Unit, University of Cambridge. The Principal Investigators are
641 Joni Holmes (Head of CALM), Susan Gathercole (Chair of CALM Management Committee),
642 Duncan Astle, Tom Manly and Rogier Kievit. Data collection is assisted by a team of
643 researchers and PhD students at the CBU that includes Sarah Bishop, Annie Bryant, Sally
644 Butterfield, Fanchea Daily, Laura Forde, Erin Hawkins, Sinead O'Brien, Clíodhna O'Leary,
645 Joseph Rennie, and Mengya Zhang. The authors wish to thank the many professionals
646 working in children's services in the South-East and East of England for their support, and to
647 the children and their families for giving up their time to visit the clinic. We would also like to
648 thank all NKI-RS participants and researchers. We are grateful to Amber Ruigrok for helpful
649 suggestions regarding pubertal development.

650

References

651 Ackerman PL, Beier ME, Boyle MO. 2005. Working memory and intelligence: The same or
652 different constructs? *Psychol Bull.* 131:30–60.
653 Alloway TP. 2007. *Automated Working Memory Assessment*. London, UK: The Psychological
654 Corporation.
655 Bathelt J, Johnson A, Zhang M, Astle DE. 2018. Data-driven brain-types and their cognitive
656 consequences. *bioRxiv*.
657 Blum D, Holling H. 2017. Spearman's law of diminishing returns. A meta-analysis. *Intelligence.*
658 65:60–66.
659 Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T,
660 Mehta P, Fox J. 2011. OpenMx: An open source extended Structural Equation
661 Modeling framework. *Psychometrika.* 76:306–317.

- 662 Brandmaier AM, von Oertzen T, McArdle JJ, Lindenberger U. 2013. Structural Equation Model
663 Trees. *Psychol Methods*. 18:71–86.
- 664 Byrne ML, Whittle S, Vijayakumar N, Dennison M, Simmons JG, Allen NB. 2017. A systematic
665 review of adrenarche as a sensitive period in neurobiological development and
666 mental health. *Sensitive Periods Dev*. 25:12–28.
- 667 Cannon TD, Keller MC. 2006. Endophenotypes in the genetic analyses of mental disorders.
668 *Annu Rev Clin Psychol*. 2:267–290.
- 669 Chevalier N, Kurth S, Doucette MR, Wiseheart M, Deoni SCL, Dean DC III, O’Muircheartaigh J,
670 Blackwell KA, Munakata Y, LeBourgeois MK. 2015. Myelination is associated with
671 processing speed in early childhood: Preliminary insights. *PLOS ONE*. 10:e0139897.
- 672 Chuderski A. 2013. When are fluid intelligence and working memory isomorphic and when
673 are they not? *Intelligence*. 41:244–262.
- 674 Cox SR, Ritchie SJ, Tucker-Drob EM, Liewald DC, Hagenaars SP, Davies G, Wardlaw JM, Gale
675 CR, Bastin ME, Deary IJ. 2016. Ageing and brain white matter structure in 3,513 UK
676 Biobank participants. *Nat Commun*. 7:13629.
- 677 Coyle TR, Pillow DR, Snyder AC, Kochunov P. 2011. Processing speed mediates the
678 development of general intelligence (g) in adolescence. *Psychol Sci*. 22:1265–1269.
- 679 de Mooij SMM, Henson RNA, Waldorp LJ, Kievit RA. 2018. Age differentiation within grey
680 matter, white matter and between memory and white matter in an adult lifespan
681 cohort. *J Neurosci*.
- 682 Delis DC, Kramer JH, Kaplan E, Holdnack J. 2004. Reliability and validity of the Delis-Kaplan
683 Executive Function System: An update. *J Int Neuropsychol Soc*. 10:301–303.
- 684 Dorn LD. 2006. Measuring puberty. *J Adolesc Health*. 39:625–626.
- 685 Dunning DL, Holmes J, Gathercole SE. 2013. Does working memory training lead to
686 generalized improvements in children with low working memory? A randomized
687 controlled trial. *Dev Sci*. 16:915–925.
- 688 Fedorenko E, Duncan J, Kanwisher N. 2013. Broad domain generality in focal regions of
689 frontal and parietal cortex. *Proc Natl Acad Sci*. 201315235.
- 690 Ferrer E, Whitaker K, Steele J, Green C, Wendelken C, Bunge S. 2013. White matter
691 maturation supports the development of reasoning ability through its influence on
692 processing speed. *Dev Sci*. 16:941–951.
- 693 Finkel D, Reynolds CA, McArdle JJ, Pedersen NL. 2005. The longitudinal relationship between
694 processing speed and cognitive ability: Genetic and environmental influences. *Behav
695 Genet*. 35:535–549.
- 696 Fry AF, Hale S. 2000. Relationships among processing speed, working memory, and fluid
697 intelligence in children. *Biol Psychol*. 54:1–34.
- 698 Fukuda K, Vogel E, Mayr U, Awh E. 2010. Quantity not quality: The relationship between fluid
699 intelligence and working memory capacity. *Psychon Bull Rev*. 17:673–679.
- 700 Gale CR, Batty GD, Tynelius P, Deary IJ, Rasmussen F. 2010. Intelligence in early adulthood
701 and subsequent hospitalisation and admission rates for the whole range of mental
702 disorders: Longitudinal study of 1,049,663 men. *Epidemiol Camb Mass*. 21:70–77.

- 703 Gallagher A, Frederickson N. 1995. The Phonological Assessment Battery (PhAB): An initial
704 assessment of its theoretical and practical utility. *Educ Child Psychol*.
- 705 Garrett HE. 1946. A developmental theory of intelligence. *Am Psychol*. 1:372–378.
- 706 Gignac GE, Szodorai ET. 2016. Effect size guidelines for individual differences researchers.
707 *Personal Individ Differ*. 102:74–78.
- 708 Gur RC, Ragland JD, Moberg PJ, Turner TH, Bilker WB, Kohler C, Siegel SJ, Gur RE. 2001.
709 Computerized Neurocognitive Scanning: I. methodology and validation in healthy
710 people. *Neuropsychopharmacology*. 25:766.
- 711 Holmes J, Bryant A, Gathercole SE. 2018. A transdiagnostic study of children with problems of
712 attention, learning and memory (CALM). *bioRxiv*.
- 713 Hua K, Zhang J, Wakana S, Jiang H, Li X, Reich DS, Calabresi PA, Pekar JJ, van Zijl PCM, Mori S.
714 2008. Tract probability maps in stereotaxic spaces: analyses of white matter anatomy
715 and tract-specific quantification. *NeuroImage*. 39:336–347.
- 716 Johnson MH. 2000. Functional brain development in infants: Elements of an interactive
717 specialization framework. *Child Dev*. 71:75–81.
- 718 Johnson MH. 2011. Interactive Specialization: A domain-general framework for human
719 functional brain development? *Dev Cogn Neurosci*. 1:7–21.
- 720 Jones DK, Knösche TR, Turner R. 2013. White matter integrity, fiber count, and other fallacies:
721 The do's and don'ts of diffusion MRI. *NeuroImage*. 73:239–254.
- 722 Jöreskog KG. 1999. How large can a standardized coefficient be?
- 723 Jung RE, Haier RJ. 2007. The Parieto-Frontal Integration Theory (P-FIT) of intelligence:
724 Converging neuroimaging evidence. *Behav Brain Sci*. 30:135–154.
- 725 Kail RV, Lervåg A, Hulme C. 2015. Longitudinal evidence linking processing speed to the
726 development of reasoning. *Dev Sci*. 19:1067–1074.
- 727 Kail RV, Salthouse TA. 1994. Processing speed as a mental capacity. *Acta Psychol (Amst)*.
728 86:199–225.
- 729 Kamali A, Flanders AE, Brody J, Hunter JV, Hasan KM. 2014. Tracing superior longitudinal
730 fasciculus connectivity in the human brain using high resolution diffusion tensor
731 tractography. *Brain Struct Funct*. 219:10.1007/s00429-012-0498-y.
- 732 Kaufman AS. 1975. Factor analysis of the WISC-R at 11 age levels between 61/2 and 161/2
733 years. *J Consult Clin Psychol*. 43:135–147.
- 734 Kievit R, Fuhrmann D, Borgeest G, Simpson-Kent I, Henson R. 2018. The neural determinants
735 of age-related changes in fluid intelligence: a pre-registered, longitudinal analysis in
736 UK Biobank [version 2; referees: 3 approved]. *Wellcome Open Res*. 3.
- 737 Kievit RA, Davis SW, Griffiths J, Correia MM, CamCAN, Henson RN. 2016. A watershed model
738 of individual differences in fluid intelligence. *Neuropsychologia*. 91:186–198.
- 739 Kievit RA, Davis SW, Griffiths J, Correia MM, Cam-CAN, Henson RN. 2016. A watershed model
740 of individual differences in fluid intelligence. *Neuropsychologia*. 91:186–198.
- 741 Kievit RA, Romeijn JW, Waldorp LJ, Wicherts JM, Scholte HS, Borsboom D. 2011. Mind the
742 gap: A psychometric approach to the reduction problem. *Psychol Inq*. 22:67–87.
- 743 Kline RB. 2015. Principles and practice of structural equation modeling. Guilford publications.

- 744 Knoll LJ, Fuhrmann D, Sakhardande AL, Stamp F, Speekenbrink M, Blakemore SJ. 2016. A
745 window of opportunity for cognitive training in adolescence. *Psychol Sci.* 27:1620–
746 1631.
- 747 Kyllonen PC, Christal RE. 1990. Reasoning ability is (little more than) working-memory
748 capacity?! *Intelligence.* 14:389–433.
- 749 Lövdén M, Laukka EJ, Rieckmann A, Kalpouzos G, Li T, Jonsson T, Wahlund L, Fratiglioni L,
750 Bäckman L. 2013. The dimensionality of between-person differences in white matter
751 microstructure in old age. *Hum Brain Mapp.* 34:1386–1398.
- 752 Lu PH, Lee GJ, Raven EP, Tingus K, Khoo T, Thompson PM, Bartzokis G. 2011. Age-related
753 slowing in cognitive processing speed is associated with myelin integrity in a very
754 healthy elderly sample. *J Clin Exp Neuropsychol.* 33:1059–1068.
- 755 Mackey AP, Hill SS, Stone SI, Bunge SA. 2011. Differential effects of reasoning and speed
756 training in children. *Dev Sci.* 14:582–590.
- 757 MacPherson SE, Cox SR, Dickie DA, Karama S, Starr JM, Evans AC, Bastin ME, Wardlaw JM,
758 Deary IJ. 2017. Processing speed and the relationship between Trail Making Test-B
759 performance, cortical thinning and white matter microstructure in older adults.
760 *Cortex.* 95:92–103.
- 761 Manly T, Anderson V, Nimmo-Smith I, Turner A, Watson P, Robertson IH. 2001. The
762 differential assessment of children’s attention: The Test of Everyday Attention for
763 Children (TEA-Ch), normative sample and ADHD performance. *J Child Psychol*
764 *Psychiatry.* 42:1065–1081.
- 765 Mills KL, Goddings A-L, Herting MM, Meuwese R, Blakemore S-J, Crone EA, Dahl RE, Güroğlu
766 B, Raznahan A, Sowell ER, Tamnes CK. 2016. Structural brain development between
767 childhood and adulthood: Convergence across four longitudinal samples.
768 *NeuroImage.* 141:273–281.
- 769 Mills KL, Tamnes CK. 2014. Methods and considerations for longitudinal structural brain
770 imaging analysis across development. *Dev Cogn Neurosci.* 9:172–190.
- 771 Noack H, Lövdén M, Schmiedek F. 2014. On the validity and generality of transfer effects in
772 cognitive training research. *Psychol Res.* 78:773–789.
- 773 Nooner K, Colcombe S, Tobe R, Mennes M, Benedict M, Moreno A, Panek L, Brown S, Zavitz
774 S, Li Q, Sikka S, Gutman D, Bangaru S, Schlachter RT, Kamiel S, Anwar A, Hinz C, Kaplan
775 M, Rachlin A, Adelsberg S, Cheung B, Khanuja R, Yan C, Craddock C, Calhoun V,
776 Courtney W, King M, Wood D, Cox C, Kelly C, DiMartino A, Petkova E, Reiss P, Duan N,
777 Thompsen D, Biswal B, Coffey B, Hoptman M, Javitt D, Pomara N, Sidtis J, Koplewicz H,
778 Castellanos F, Leventhal B, Milham M. 2012. The NKI-Rockland Sample: A model for
779 accelerating the pace of discovery science in psychiatry. *Front Neurosci.* 6:152.
- 780 Oberauer K, Schulze R, Wilhelm O, Süß H-M. 2005. Working memory and intelligence - Their
781 correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005).
782 *Psychol Bull.* 131:61–65.
- 783 Orben A, Przybylski AK. 2019. The association between adolescent well-being and digital
784 technology use. *Nat Hum Behav.*

- 785 Parlolini V, Radua J, Dell'Acqua F, Leslie A, Simmons A, Murphy DG, Catani M, Thiebaut de
786 Schotten M. 2017. Functional segregation and integration within fronto-parietal
787 networks. *NeuroImage*. 146:367–375.
- 788 Penke L, Maniega SM, Murray C, Gow AJ, Valdés Hernández MC, Clayden JD, Starr JM,
789 Wardlaw JM, Bastin ME, Deary IJ. 2010. A general factor of brain white matter
790 integrity predicts information processing speed in healthy older people. *J Neurosci*.
791 30:7569.
- 792 Pfefferbaum A, Sullivan EV, Hedehus M, Lim KO, Adalsteinsson E, Moseley M. 2000. Age-
793 related decline in brain white matter anisotropy measured with spatially corrected
794 echo-planar diffusion tensor imaging. *Magn Reson Med*. 44:259–268.
- 795 Primi R, Ferrão ME, Almeida LS. 2010. Fluid intelligence as a predictor of learning: A
796 longitudinal multilevel approach applied to math. *Learn Individ Differ*. 20:446–451.
- 797 R core team the. 2015. R: A language and environment for statistical computing.
- 798 Ritchie SJ, Tucker-Drob EM. 2018. How much does education improve intelligence? A meta-
799 analysis. *Psychol Sci*. 0956797618774253.
- 800 Rosseel Y. 2012. lavaan: An R package for Structural Equation Modeling. *J Stat Softw*. 48:1–
801 36.
- 802 Roth B, Becker N, Romeyke S, Schäfer S, Domnick F, Spinath FM. 2015. Intelligence and
803 school grades: A meta-analysis. *Intelligence*. 53:118–137.
- 804 Salthouse TA. 1996. The processing-speed theory of adult age differences in cognition.
805 *Psychol Rev*. 103:403–428.
- 806 Schermelleh-Engel K, Moosbrugger H, Müller H. 2003. Evaluating the fit of Structural
807 Equation Models: Tests of significance and descriptive goodness-of-fit measures.
808 *Methods Psychol Res*. 8:23–74.
- 809 Schubert A-L, Hagemann D, Frischkorn GT. 2017. Is general intelligence little more than the
810 speed of higher-order processing? *J Exp Psychol Gen*. 146:1498–1512.
- 811 Schwaighofer M, Fischer F, Bühner M. 2015. Does working memory training transfer? A
812 meta-analysis including training conditions as moderators. *Educ Psychol*. 50:138–166.
- 813 Scott KM, Kline M. 2019. Enabling Confirmatory Secondary Data Analysis by Logging Data
814 Checkout. *Adv Methods Pract Psychol Sci*. 2515245918815849.
- 815 Sheppard LD, Vernon PA. 2008. Intelligence and speed of information-processing: A review of
816 50 years of research. *Personal Individ Differ*. 44:535–551.
- 817 Simpson-Kent IL, Fuhrmann D, Bathelt J, the CALM team, Kievit RA. 2019. Cross-sectional
818 evidence for age differentiation-dedifferentiation in crystallized and fluid intelligence
819 and their white matter substrates during childhood and adolescence. Prep.
- 820 Sisk CL, Zehr JL. 2005. Pubertal hormones organize the adolescent brain and behavior. *Front*
821 *Neuroendocrinol*. 26:163–174.
- 822 Tamnes CK, Roalf DR, Goddings A-L, Lebel C. 2017. Diffusion MRI of white matter
823 microstructure development in childhood and adolescence: Methods, challenges and
824 progress. *Dev Cogn Neurosci*.

- 825 Tucker-Drob EM. 2009. Differentiation of cognitive abilities across the lifespan. *Dev Psychol.*
826 45:1097–1118.
- 827 Vestergaard M, Madsen KS, Baaré WF, Skimminge A, Ejersbo LR, Ramsøy TZ, Gerlach C,
828 Åkeson P, Paulson OB, Jernigan TL. 2011. White matter microstructure in superior
829 longitudinal fasciculus associated with spatial working memory performance in
830 children. *J Cogn Neurosci.* 23:2135–2146.
- 831 Volkow ND, Koob GF, Croyle RT, Bianchi DW, Gordon JA, Koroshetz WJ, Pérez-Stable EJ, Riley
832 WT, Bloch MH, Conway K, Deeds BG, Dowling GJ, Grant S, Howlett KD, Matochik JA,
833 Morgan GD, Murray MM, Noronha A, Spong CY, Wargo EM, Warren KR, Weiss SRB.
834 2018. The conception of the ABCD study: From substance use to a broad NIH
835 collaboration. *Adolesc Brain Cogn Dev ABCD Consort Ration Aims Assess Strategy.*
836 32:4–7.
- 837 Vollmar C, O’Muircheartaigh J, Barker GJ, Symms MR, Thompson P, Kumari V, Duncan JS,
838 Richardson MP, Koepp MJ. 2010. Identical, but not the same: intra-site and inter-site
839 reproducibility of fractional anisotropy measures on two 3.0T scanners. *NeuroImage.*
840 51:1384–1394.
- 841 Wagenmakers E-J, Farrell S. 2004. AIC model selection using Akaike weights. *Psychon Bull*
842 *Rev.* 11:192–196.
- 843 Wagenmakers E-J, Wetzels R, Borsboom D, van der Maas HLJ, Kievit RA. 2012. An agenda for
844 purely confirmatory research. *Perspect Psychol Sci.* 7:632–638.
- 845 Wandell BA. 2016. Clarifying human white matter. *Annu Rev Neurosci.* 39:103–128.
- 846 Wechsler D. 2011. *Wechsler Abbreviated Scale of Intelligence -Second Edition.* San Antonio,
847 TX: Pearson.
- 848 Weston SJ, Ritchie SJ, Rohrer JM, Przybylski AK. 2018. Recommendations for increasing the
849 transparency of analysis of pre-existing datasets. *PsyArXiv.*
- 850