

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 SEM, as a statistical technique, is uniquely suited to modeling the kinds of complex
104 multivariate brain-behavior associations posited by the watershed model (Kievit et al. 2011;
105 Kline 2015). SEM combines factor analysis and path analysis (a variant of regression analysis).

106 It can model abstract cognitive constructs like g_f , by estimating latent variables from
107 observed task scores (i.e. manifest variables). This feature of SEM allowed us to model g_f ,
108 working memory, and processing speed in two independent samples, and thereby provided a
109 direct test of the generalizability of our findings. Second, SEM can test the simultaneous
110 relations between multiple cognitive and neural variables, allowing us to address the relative
111 contributions of different white matter tracts and different cognitive endophenotypes to g_f .
112 Finally, using SEM Trees (Brandmaier et al. 2013), a novel, decision-tree-based extension of
113 SEM, we investigated whether the associations in the watershed model change with age.

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

- 116 1. Working memory, g_f and processing speed are separable constructs.
- 117 2. Individual differences in g_f are predicted by working memory and processing speed.
- 118 3. White matter microstructure is a multi-dimensional construct.
- 119 4. There is a hierarchical relationship between white matter microstructure, cognitive
120 endophenotypes (working memory and processing speed) and g_f , such that white

121 matter contributes to working memory and processing speed, which, in turn
122 contribute to g_f .

123 5. The contribution of working memory and processing speed to g_f changes with age.

124 **Materials and Methods**

125 **Samples**

126 We analysed data from the CALM and NKI-RS sample, as described in detail by (Holmes et al.
127 2018) and (Nooner et al. 2012) respectively. See also Simpson-Kent et al. (2019). We had also
128 preregistered to analyse data from the ABCD cohort (Volkow et al. 2018). The latter cohort
129 contains only data for 9 - and 10 - year olds at present, however, which limits comparability
130 to CALM and NKI-RS, and makes it unsuitable for investigations of developmental differences.
131 We therefore opted to not analyse ABCD data here and instead recommend a replication of
132 the analyses presented here in ABCD once longitudinal data is available. The CALM sample
133 consists of children and adolescents referred by health and educational professionals as
134 having difficulties in attention, learning and/or memory. The NKI-RS is a community-
135 ascertained, lifespan sample, and representative of the general population of Rockland, New
136 York, and the United States as a whole, in terms of ethnicity, socioeconomic status etc. For
137 NKI-RS, we included data for participants under the age of 18 only to match the age range of
138 CALM and excluded data that were completed more than half a year after enrolment. The
139 latter criterion was implemented to ensure that age at assessment did not differ
140 substantively between cognitive measures. The final samples included 551 participants from
141 CALM (30.85% female, aged 5.17 - 17.92 years, $N_{\text{Neuroimaging}} = 165$) and 335 participants from
142 NKI-RS (43.48% female, aged 6.06 - 17.92 years, $N_{\text{Neuroimaging}} = 67$). See Table 1 for prevalence
143 of relevant disorders and learning difficulties in the samples.

144 *Table 1. Prevalence of Relevant Disorders and Learning Difficulties in the CALM and NKI-RS*
 145 *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

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

147 **Cognitive Tasks**

148 We included cognitive tasks measuring the domains of g_f , working memory or processing
 149 speed for CALM and NKI-RS. See Table 2 for the complete list of tasks used, and the
 150 Supplementary Methods for task descriptions. Supplementary Figure 1 and 2 show raw
 151 scores on all tasks. The tasks modelled here were preregistered for CALM but not NKI-RS.

152 *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

153 *Note.* See the Supplementary Methods for task descriptions. Abbreviations: AWMA -
 154 Automated Working Memory Assessment (Alloway 2007), CNB - Computerized
 155 Neurocognitive Battery (Gur et al. 2001), DKEF - Delis-Kaplan Executive Functioning System
 156 (Delis et al. 2004), PhAB – Phonological Assessment Battery (Gallagher and Frederickson
 157 1995), TEA-Ch RBBS - Test of Everyday Attention for Children, Red & Blues, Bags & Shoes
 158 subscale (Manly et al. 2001), WASI - Wechsler Abbreviated Scale of Intelligence - Second
 159 Edition (Wechsler 2011), WISC-R - Wechsler Intelligence Scale for Children – Revised
 160 (Kaufman 1975).

161 **White Matter Microstructure**

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

166 **Analysis Methods and Structural Equation Modelling**

167 Covariance matrices and scripts replicating key analyses can be obtained from:
 168 https://github.com/df1234/gf_development. Supplementary Figure 5 and 6 show correlation
 169 matrices of all tasks and white matter tracts modelled. We modelled raw scores for g_f and

170 working memory tasks, as preregistered. Raw scores on processing speed tasks were
171 transformed. This step was not preregistered, but found necessary to achieve model
172 convergence to ensure interpretability of scores. First, we inverted response time scores
173 (using the formula $y = 1/x$) to obtain more intuitive measures of 'speed' for all but the CNB
174 Motor Speed task, for which raw scores were already a measure of speed. Afterwards, we
175 applied a log-transformation to reaction time tasks to increase normality and aid estimation.
176 For the CNB Motor Speed task only, we additionally removed values $\pm 2 SD$ of the mean ($N =$
177 6) because the presence of these outliers had caused convergence problems.

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

182 We used SEM Trees to investigate whether the associations among cognitive and neural
183 measures differed with age. SEM Trees use decision tree methods to hierarchically split a
184 dataset into subgroups if parameter estimates differ significantly based on a covariate of
185 interest - in this case age (Brandmaier et al. 2013). We first ran the watershed model in
186 OpenMx (Boker et al. 2011) and then passed this model object to semtree to compute the
187 SEM Trees. We ran one SEM Tree for each parameter of interest (e.g. the covariance
188 between working memory and processing speed). All other parameters in each semtree
189 object were set to be invariant across groups to ensure that splits were specific to the
190 parameter of interest. We used a 10 - fold cross-validation estimation method as recommend
191 by (Brandmaier et al. 2013). For the path from the cingulate to working memory only we
192 used 5 - fold cross-validation because the model did not converge using 10 - fold cross-

193 validation. Minimum sample size in age group was set to $N = 50$ to ensure reliable estimation
194 of standard errors. Note that this choice effectively limited search space for potential splits to
195 ages 6.58 - 12.42 years for CALM and 8.08 - 15.49 years for NKI-RS.

196 **Results**

197 To evaluate the hypotheses generated by the watershed model, we built up the watershed
198 model in steps and carried our comprehensive tests of model fit at each step. First, we
199 assessed the overall fit of our models to the data using the chi-square test, root mean square
200 error of approximation (RMSEA), comparative fit index (CFI) and standardized root mean
201 square residual (SRMR). Good absolute fit was defined as $RMSEA < 0.05$, $CFI > 0.97$ and $SRMR$
202 < 0.05 ; acceptable fit as $RMSEA = 0.08 - 0.05$, $CFI = 0.95 - 0.97$, $SRMR = 0.05 - 0.10$
203 (Schermelleh-Engel et al. 2003). Second, we assessed specific predictions from our models by
204 comparing them to alternative models. Comparative model fit for nested models was
205 assessed using the chi-square difference test. Non-nested models were compared using the
206 Akaike (AIC) weights, which indicates the probability of a model being the data-generating
207 model compared to all other models tested (Wagenmakers and Farrell 2004). Lastly, we
208 evaluated the significance and strength of relationships between specific variables in our
209 models by inspecting the Wald test for individual parameters, noting the joint R^2 where
210 relevant and reporting standardized parameter estimates. Absolute standardized parameter
211 estimates above 0.10 were defined as small effects, 0.20 as typical and 0.30 as large (Gignac
212 and Szodorai 2016).

213 The Measurement Model of Cognition

214 To examine the neurocognitive architecture of g_f , we started by modelling the cognitive
215 components of the watershed model: g_f , working memory and processing speed. Specifically,
216 we fit a three-factor model of cognition (Figure 3) and compared it to alternative
217 measurement models. This approach allowed us to test Hypothesis 1: namely that g_f , working
218 memory and processing speed form three separable, albeit likely correlated cognitive factors.

219 The Three-Factor Model (Figure 3) showed excellent *absolute fit* for both the CALM and NKI-
220 RS sample (Table 3), indicating that overall, the data was compatible with a model of g_f ,
221 working memory and processing speed as three separate factors.

222 The Three-Factor Model also showed very good *comparative fit* for NKI-RS as well, with a
223 96.60% probability of being the data-generating model compared to all alternative models
224 tested, as indicated by its AIC weight (Figure 3). The evidence was more mixed for CALM, for
225 which the Three-Factor Model showed a 27.15% probability of being the data-generating
226 model, while Two-Factor Model B (Figure 3, treating working memory and g_f as a unitary
227 factor) showed a 72.85% probability of being the data-generating model, highlighting a close
228 relationship between g_f and working memory for this sample. The Single-Factor Model and
229 Two-Factor Model A (Figure 3, treating speed and g_f as a unitary factor) showed a very low
230 (approximately 0%) probability of being the data-generating model, indicating that speed and
231 g_f were clearly separable in both samples.

232 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%

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

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

243 *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, $z = 17.43$, $p < .001$

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

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

246 We next examined the relationships between working memory, processing speed and g_f in
 247 more detail. Specifically, we fit a SEM including regression paths between working memory
 248 and g_f , as well as speed and g_f , to test Hypothesis 2 - that working memory and processing
 249 speed each predict individual differences in g_f . We found that this model showed good
 250 absolute fit for both samples (CALM: $\chi^2(18) = 41.74$, $p = .001$; RMSEA = .049 [.030 - .068]; CFI
 251 = .983; SRMR = .032, NKI-RS: $\chi^2(32) = 54.15$, $p = .009$; RMSEA = .045 [.024 - .065]; CFI = .981;
 252 SRMR = .030), indicating that, overall, the data was compatible with our model.

253 To further scrutinize the relationship between g_f , working memory and speed, we compared
 254 our freely-estimated model to a set of alternative models with different constraints imposed
 255 upon the regression paths. First, to test whether working memory and speed each made
 256 different contributions, we tested an alternative model in which the paths from processing
 257 speed and working memory to g_f were constrained to be equal. In CALM ($\Delta\chi^2(1) = 15.53$, $p <$
 258 $.001$), but not NKI-RS ($\Delta\chi^2(1) = 3.25$, $p = .072$), the freely-estimated model fit better than the
 259 equality-constrained model, indicating that working memory and speed each made different
 260 contributions in CALM but not NKI-RS. Next, we tested whether the freely estimated model
 261 fit better than a model in which the path between g_f and working memory was constrained

262 to zero. We found that that the freely estimated model fit better for both samples (CALM:
 263 $\Delta\chi^2(1) = 20.77, p < .001$; NKI-RS: $\Delta\chi^2(1) = 12.97, p < .001$). In line with our hypothesis, this
 264 result indicates that working memory makes a significant incremental contribution to g_f .
 265 Finally, we tested a model in which the path between g_f and processing speed was
 266 constrained to zero. This model showed no difference in fit to the freely estimated model for
 267 CALM ($\Delta\chi^2(1) = 0.02, p = .875$) or NKI-RS ($\Delta\chi^2(1) = 0.04, p = .849$). Contrary to our hypothesis,
 268 this indicates that there was no clear incremental contribution of processing speed to g_f .
 269 Finally, we inspected standardized path estimates of the freely estimated model to assess the
 270 effect sizes of working memory and processing speed. Parameter estimates showed that
 271 working memory showed a greater effect on g_f than processing speed, particularly in CALM
 272 (Table 5) even though raw correlations between g_f and speed were high in both samples
 273 (Table 4).

274 *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, $z = 1.81, p = .070$

275
 276 Overall these results provide mixed evidence for Hypothesis 2: There was good evidence that
 277 working memory and speed made a significant joint contribution to g_f , and that working
 278 memory made an incremental contribution to g_f in CALM. Contrary to our hypothesis, and
 279 the watershed model, however, processing speed showed no significant incremental
 280 contribution to g_f , above and beyond working memory. We explore likely explanations for this
 281 finding in the Discussion.

282 The Measurement Model of White Matter

283 We next examined the measurement model of white matter to test Hypothesis 3, namely
284 that white matter microstructure is a multi-dimensional construct. Specifically, we examined
285 absolute model fit of a single factor model to test whether a unidimensional model could
286 adequately capture white matter microstructure. As expected, the single-factor model of
287 white matter microstructure did not fit the data well (CALM: $\chi^2(35) = 124.63$, $p < .001$;
288 RMSEA = .125 [.103 - .147]; CFI = .933; SRMR = .039; NKI-RS: $\chi^2(35) = 132.33$, $p < .001$; RMSEA
289 = .204 [.167 - .242]; CFI = .885; SRMR = .023). This indicates that white matter microstructure
290 could not be reduced to a single 'global FA' dimension in our samples, in line with (Lövdén et
291 al. 2013; Kievit, Davis, Griffiths, Correia, CamCAN, et al. 2016) and supporting Hypothesis 3.
292 We therefore modelled each of the ten white matter tracts separately in all subsequent
293 models.

294 The Watershed Model: Relationships between Cognition and White Matter

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

301 We found largely converging results across samples. The watershed model showed good
302 absolute fit in CALM ($\chi^2(78) = 107.78$, $p = .014$; RMSEA = .026 [.012 - .038]; CFI = .981; SRMR
303 = .043) and acceptable fit in NKI-RS ($\chi^2(112) = 219.22$, $p < .001$; RMSEA = .053 [.043 - .064];
304 CFI = .928; SRMR = .088). White matter explained large amounts of variance in working

305 memory ($R^2_{\text{CALM}} = 32.3\%$; $R^2_{\text{NKI-RS}} = 46.1\%$) and processing speed ($R^2_{\text{CALM}} = 38.2\%$; $R^2_{\text{NKI-RS}} =$
306 54.4%), which, in turn, explained even more variance in g_f ($R^2_{\text{CALM}} = 51.2\%$; $R^2_{\text{NKI-RS}} = 78.3\%$).

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

308 Comparing the freely estimated watershed model to alternative, constrained, models
309 showed that white matter contributed significantly to memory and processing speed.

310 Specifically, a model in which paths from white matter to processing speed were constrained
311 to zero fit worse than the freely-estimated model (CALM: $\Delta\chi^2(10) = 50.26$, $p < .001$; NKI-RS:

312 $\Delta\chi^2(10) = 27.19$, $p = .002$), as did a model in which paths from white matter to working

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

314 $p = .004$). As hypothesised, white matter therefore contributed to both processing speed and

315 working memory.

316 We next inspected that relationship between individual white matter tracts and working
317 memory and speed in more detail. A model in which paths from white matter to working

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

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

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

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

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

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

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

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

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

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

328 processing speed and working memory (Figure 5). Two tracts showed an unexpected,
329 strongly negative (< -1), relationship: the forceps minor for CALM and the inferior fronto-
330 occipital fasciculus for NKI-RS. We found that these negative estimates occurred only when
331 all other brain to cognition pathways were also estimated: When estimated on their own,
332 path estimates were positive (forceps minor to working memory: standardized estimate =
333 0.36, $z = 4.05$, $p < .001$; inferior fronto-occipital fasciculus to working memory: standardized
334 estimate = 0.14, $z = 0.859$, $p = .390$; inferior fronto-occipital fasciculus to processing speed:
335 standardized estimate = 0.26, $z = 1.41$, $p = .158$). This sign-flip suggests that the negative
336 pathways were potentially due to modelling several, highly-correlated paths at the same time
337 (Jöreskog 1999). Overall, these results further support the watershed prediction that multiple
338 white matter tracts map onto working memory and processing speed.

339 Finally, we probed the watershed model in more detail by testing a set of alternative
340 expressions of the watershed model still compatible with the core tenants of the watershed
341 model – as well as a set of alternative models incompatible with the watershed model. We
342 compared all alternatives (see Figure 6 for graphical representations) to the original
343 watershed model by inspecting each models' relative probability of being the data-generating
344 model as indicated by AIC weights (Wagenmakers and Farrell 2004). We found that the
345 original watershed model showed a very high probability (98.58%) of being the data-
346 generating model for CALM but only a 0.10% probability for NKI-RS. For NKI-RS, a different
347 expression of the watershed model, such that g_f was regressed on working memory, which
348 was regressed on processing speed, which was then regressed on white matter (Alternative
349 A, Figure 6) showed a 95.04% probability of being the data-generating model. This model
350 only showed a 0.37% probability for CALM. Another expression of the watershed model, in
351 which all tasks were modelled separately as manifest, rather than latent, variables

352 (Alternative B, Figure 6), showed no advantage over the watershed model for CALM (0.00%
353 probability) or NKI-RS (0.00% probability). We next tested two alternative models
354 incompatible with the tenants of the watershed model. We found that a model in which the
355 hierarchy between cognitive endophenotypes and g_f was inverted (Alternative C, Figure 6)
356 showed comparatively low probability of being the data-generating model for both CALM
357 (0.00%) and NKI-RS (2.86%). Similarly, a model in which g_f was directly regressed on white
358 matter, working memory and processing speed (Alternative D, Figure 6), showed no clear
359 advantage over the watershed model for CALM (1.05% probability) or NKI-RS (0.00%
360 probability). Overall these model comparisons highlight that while the watershed model fit
361 the data for both samples and had large explanatory power (as indicated by R^2 s), the precise
362 configuration of the watershed model may differ somewhat between cohorts.

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

371 **Testing for potential confounds**

372 We carried out a series of supplementary and non-preregistered analyses to examine
373 whether possible confounders influenced our models. These analyses showed that our
374 findings were robust to the inclusion of covariates such as scanner motion or socio-economic

375 status. They were also robust across genders and participants taking or not taking
376 medication. There were no differences in the structure of the model between participants
377 with and without diagnosed disorders for CALM. Potential small differences cannot be ruled
378 out for NKI-RS, likely due to the low number of diagnosed participants of $N = 106$
379 (Supplementary Material).

380 **Age-Related Differences in the Neurocognitive Architecture of g_f**

381 Finally, we tested Hypothesis 5 - that the contribution of working memory and processing
382 speed to g_f varied with age. We first inspected cross-sectional differences in g_f , working
383 memory and processing speed, and then used SEM trees to investigate potential age-
384 differences in the relationships between these factors. In addition, non-preregistered,
385 analyses we also used SEM Trees to investigate potential age-differences in the relationship
386 between white matter and cognitive endophenotypes by inspecting paths that were
387 significant in the watershed model (Figure 4 and 5).

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

397 *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	-	-

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

404 As shown in Figure 7, g_f , working memory and processing speed factor scores increased with
 405 age for all three cognitive phenotypes. In line with our hypothesis, SEM trees showed that
 406 there were pronounced age-related differences in brain-behaviour in childhood and
 407 adolescence (Table 6). For both samples and all but one path, there was an initially strong
 408 relationship between components of the watershed model, then a dip around ages 7 - 9
 409 years for CALM and age 8 for NKI-RS, followed by an increase in path strength around ages 11
 410 - 12 (see Supplementary Figure 7 for a graphical representation of these results).
 411 Speculatively, this pattern of results is consistent with an interpretation of a reorganization of

412 neurocognitive faculties in late childhood, followed by a consolidation of neurocognitive
413 pathways around the onset of adolescence (Johnson 2000, 2011).

414 Discussion

415 We here used multivariate statistical techniques to investigate the neurocognitive
416 architecture of g_f in two large ($N_{\text{CALM}} = 551$, $N_{\text{NKI-RS}} = 335$) developmental cohorts and, for the
417 first time, investigated how the neurocognitive architecture of g_f changes dynamically with
418 age. We tested a preregistered watershed model of g_f , which predicts a hierarchy of partially
419 independent effects. As might be expected from a multi-cohort study, there were some
420 differences between the community-ascertained cohort (NKI-RS) and the cohort of children
421 and adolescents with learning difficulties (CALM) in specific path estimates. Overall however,
422 we found convergent results across these two heterogeneous samples. The watershed model
423 performed well for both CALM and NKI-RS: White matter contributed to working memory
424 and processing speed, which, in turn, contributed to g_f and explained 51% of variance therein
425 for the CALM sample and 78% of variance for NKI-RS. Models were robust across genders,
426 participants taking or not taking medication and when controlling for socio-economic status
427 and scanner motion. Investigations of age effects showed that the relationship between
428 cognitive abilities and white matter dipped in strength around ages 7-12 years. Speculatively,
429 this age-effect may reflect a reorganization of the neurocognitive architecture during pre-
430 puberty and early puberty (Byrne et al. 2017). These findings have implications for
431 understanding and targeting cognitive impairments in populations with learning difficulties.

432 The watershed model tested here consists of three levels: g_f forms the most down-stream
433 point, with working memory and processing speed as intermediate tributaries, and white

434 matter microstructural tracts as upstream sources. Previous studies suggested that matter
435 microstructure is best characterised by a single, 'global FA' factor (Penke et al. 2010) while
436 others have contended that association patterns among different white matter tracts are
437 more complex (Lövdén et al. 2013; Kievit, Davis, Griffiths, Correia, CamCAN, et al. 2016). Here
438 we found strong evidence for a multifactorial view of white matter tracts – for both samples,
439 a unidimensional model of white matter fit poorly, and for CALM, multiple tracts also showed
440 partially independent contributions to distal cognitive outcomes. This is in line with the
441 watershed model. There were some differences between cohorts as to which tracts
442 contributed most to working memory and processing speed: In line with previous research
443 (Kievit, Davis, Griffiths, Correia, CamCAN, et al. 2016; MacPherson et al. 2017; Bathelt et al.
444 2018), we found that the anterior thalamic radiation was related to processing speed, as
445 were the forceps major, forceps minor and the cingulum to working memory for CALM.
446 However, these tracts were not significant for NKI-RS. A possible explanation for these
447 differences between samples is the discrepancy in the number of participants with imaging
448 data ($N = 165$ in CALM versus $N = 67$ in NKI-RS). This discrepancy likely confers differential
449 power to detect weaker pathways. Other, not mutually exclusive, explanations are that the
450 observed differences reflect differences in brain-behaviour mapping between more atypical
451 and typical cohorts (Bathelt et al. 2018), sampling variance across two independent cohorts
452 collected under somewhat different socio-economic conditions (United Kingdom and United
453 States of America), or a more uniform age distribution in NKI-RS. While DTI images were
454 processed with the same pipeline across sites, the scanner and MRI acquisition protocol were
455 also different. Although previous work suggests that FA is relatively robust measure in multi-
456 site comparisons (Vollmar et al. 2010), we therefore cannot rule out site differences as a
457 potential confound. It will therefore be necessary to replicate these findings in large typical

458 and atypical cohorts collected in the same setting. Of note, however, the superior
459 longitudinal fasciculus was robustly associated with working memory across the two different
460 samples and settings. For NKI-RS, the superior longitudinal fasciculus was also associated with
461 processing speed. The superior longitudinal fasciculus is a large, bilateral association fibre
462 connecting temporal, occipital, parietal and frontal regions (Kamali et al. 2014). It is therefore
463 well-situated for supporting cognitive processes such as g_f , which rely on integrative multiple-
464 demand systems (Jung and Haier 2007; Fedorenko et al. 2013; Parlatini et al. 2017).

465 Our findings for the cognitive levels of the watershed model highlighted a close relationship
466 between working memory and g_f . Previous studies had variably suggested that g_f and working
467 memory (Kyllonen and Christal 1990; Fukuda et al. 2010), or g_f and processing speed (Kail and
468 Salthouse 1994; Salthouse 1996; Coyle et al. 2011; Ferrer et al. 2013) may be most closely
469 related. We found that all three cognitive factors were highly correlated for both samples.
470 Nonetheless, processing speed formed a cognitive factor clearly separable from working
471 memory and g_f . Working memory and g_f , in turn, were separable in the community-
472 ascertained NKI-RS but not in CALM, the cohort of children and adolescents with learning
473 difficulties. This close relationship between g_f and working memory was also evident in other
474 models of CALM where processing speed and working memory were used as joint predictors
475 of g_f . Contrary to our hypotheses, processing speed became non-significant after controlling
476 for working memory here. There are several possible, and not mutually exclusive,
477 explanations for this finding and the apparent differences between cohorts. First, a broader
478 set of speed tasks (which might be captured by several latent variables for clerical speed,
479 choice reaction time and speed variability) might show higher predictive power than the
480 single latent variable for speed, which could be modelled here. This may be particularly

481 pertinent for heterogeneous cohorts like CALM. Second, and in line with previous work
482 showing that time-constraints increase isomorphism of g_f and working memory (Chuderski
483 2013), even standard implementations of g_f tasks may place considerable time-pressure on
484 struggling learners, thereby increasing g_f - working memory covariance in CALM as compared
485 to NKI-RS. Conversely, less subjective or objective time pressure may also confer a
486 differentiation of cognitive domains and the watershed hierarchy in cohorts of older ages
487 and/or higher ability levels, such as NKI-RS: There was some evidence that, for this sample,
488 speed formed an intermittent level in the hierarchy between white matter and working
489 memory (Alternative A, Figure 6). Future longitudinal research will be necessary to
490 differentiate these different configurations of the watershed model and scrutinize the causal
491 flow of effects. For now, our findings highlight the value of replicating analyses in different
492 cohorts using different tasks: While evidence was mixed for the association between g_f and
493 processing speed, the strong associations between g_f and working memory across samples
494 indicate a robust and likely generalizable relationship between these two domains,
495 supporting the notion that mental information processing capacity is a key determinant of g_f
496 (Kyllonen and Christal 1990; Fukuda et al. 2010).

497 The associations in the watershed model differed between ages in a complex, non-monotonic
498 fashion. Previous research had suggested either a decrease in covariance among cognitive
499 domains with age (age differentiation; Garrett 1946), an increase in covariance with age (age
500 de-differentiation; Blum and Holling 2017), or no changes with age (Tucker-Drob 2009; de
501 Mooij et al. 2018). These investigations have traditionally focussed on relations between
502 cognitive domains, however, not on relationships between brain and cognition - although
503 see de Mooij et al. (2018). Possible linear and non-linear changes in brain-behaviour mapping

504 with age have remained mostly unexplored (Tamnes et al. 2017). Using structural equation
505 modelling trees, a novel decision-tree-based technique, we here found evidence of complex
506 developmental differences consistent across samples and relationships in the watershed
507 model: Initially strong path estimates showed a pronounced decrease in strength around
508 ages 7 - 9 years, followed by a renewed increase in the strength, even surpassing initial levels,
509 around ages 10 - 15.

510 There are at least two possible explanations for this developmental dip in brain-cognition
511 relationships. First, there may be a true decrease in relationship strength during this time of
512 life. Possibly, other cognitive skills, such as verbal reasoning, temporarily support g_f , resulting
513 in weaker relationships between g_f and working memory. Alternatively, the configuration of
514 the watershed model may change temporarily during this time, which could also manifest in
515 an apparently weaker covariance structure. In this case, the true relationship between g_f ,
516 memory, speed and white matter may still be strong, just configured differently from the
517 watershed model. We note that both explanations are compatible with the interactive
518 specialization theory (Johnson 2000, 2011), which predicts a remapping of the relationships
519 between brain substrates and cognitive abilities during development.

520 On a physiological level, this age effect may be driven by neuroendocrine changes during pre-
521 and early puberty. Puberty is driven by a complex and only partially understood set of
522 hormonal events including gonadarche and adrenarche (Sisk and Zehr 2005). Gonadarche
523 begins with the secretion of gonadotropin-releasing hormone from the hypothalamus around
524 ages 10-11 years and closely tracks the overt bodily changes of puberty (Dorn 2006).
525 Adrenarche, beginning with the maturation of the adrenal gland, starts as early as six years
526 of age, and is increasingly recognized as a complimentary driver of puberty and brain

527 development (Byrne et al. 2017). It is possible that the hormonal changes of adrenarche and
528 early gonadarche may lead to a level of neural reorganization, which may initially appear as
529 weaker relationships in the watershed model. The sweeping bodily, social and cognitive
530 changes happening in early adolescence may then drive a consolidation of the neurocognitive
531 architecture of g_r .

532 On a more general level, these age effects suggest the existence of potential non-linear
533 changes in brain-behaviour mapping during childhood and adolescence and underline the
534 value of modern statistical approaches, such as SEM Trees, for the study of age-related
535 differences. It is worth noting, however, that these findings, which are based on an inherently
536 exploratory technique, will need to be replicated in future confirmatory studies with fine-
537 grained data on puberty and larger sample sizes. The latter will also allow for detailed
538 investigations of potential gender differences. Moreover, while we were able to investigate
539 *individual differences* in g_r , we could not assess *intra-individual changes* during childhood and
540 adolescence. Although the relatively narrow age range makes large cohort effects unlikely, it
541 may still be that there were differences in recruitment and selection that varied across the
542 age range. As such, the cross-sectional nature of our samples limits our ability to make
543 inferences about developmental dynamics.

544 Our study illustrates some of the advantages and challenges of preregistered secondary data
545 analyses. We agree with others in the field that secondary data analysis need not be and
546 should not be confounded with purely exploratory research (Mills and Tamnes 2014; Orben
547 and Przybylski 2019; Scott and Kline 2019). Preregistrations, as well as dedicated multivariate
548 methods such as SEM, can help to reduce the scope for analytic flexibility and increase
549 scientific rigour when using rich, secondary datasets. Preregistrations also do not preclude

550 the use of exploratory methods or the ability to ask exploratory questions, as we did in our
551 analysis of age effects. Preregistrations merely facilitate the distinction between exploratory
552 and confirmatory research (Wagenmakers et al. 2012). There are, however, some unique
553 challenges to preregistering secondary data analyses worth noting. First, information on the
554 precise measures collected is not always available prior to data access, which can limit the
555 level of detail in which an analysis can be preregistered. Second, data quality and the level of
556 data-processing, the latter being particularly relevant for MRI data, is not always clear a priori
557 (e.g. see Kievit et al. 2018), which can necessitate changes to analyses plans after data
558 inspection. Third, convergence issues are fairly common when using complex multivariate
559 methods such as SEM. We found it necessary to transform some of our speed variables, for
560 instance, to achieve model convergence. Such post-hoc modifications, not guided by the
561 palatability of the results, but rather by unforeseen, and sometimes unforeseeable, practical
562 considerations, mean that preregistration can sometimes fall short of full compliance.
563 Nevertheless, we believe that even imperfect preregistrations, alongside shared code, data
564 and the transparent presentation of results, can help the reader distinguish between
565 confirmatory and exploratory results, and adjust their level of confidence in conclusions
566 accordingly. For guidance on maximizing transparency in preregistration of secondary data,
567 see Weston et al. (2018).

568 Finally, the findings from our study have implications understanding and targeting cognitive
569 impairments in populations with learning difficulties. First, the close relationship between
570 working memory and g_f found here and in other studies (Fukuda et al. 2010; Chuderski 2013),
571 indicates that children and adolescents struggling with working memory are likely to also
572 struggle in terms of complex reasoning tasks. Either reducing working memory load,

573 decreasing time constraints, or training working memory and fluid ability capacity in such
574 populations may therefore be promising lines of inquiry for intervention studies. It is worth
575 highlighting, however, that cognitive training studies have so far shown little evidence of (far)
576 transfer: Training abstract reasoning, a common measure of g_f , has not resulted in robust
577 increases in working memory (Knoll et al. 2016) and working memory training has not been
578 shown to transfer to reasoning skills or school performance (Dunning et al. 2013;
579 Schwaighofer et al. 2015). Similarly, transfer from processing speed to reasoning seems to be
580 limited (Mackey et al. 2011). The results obtained here suggest that interventions may
581 increase their chance of success by implementing programs of sufficient complexity to affect
582 the entire neurocognitive architecture of effects (see also Kievit et al. 2016). The level of
583 intensity required to produce sustained benefits may need to be as demanding and
584 consistent as education itself, which shows robust effects in increasing general cognitive
585 abilities over time (Ritchie and Tucker-Drob 2018). This work and work by others (Noack et al.
586 2014) also highlights the value of assessing, modeling, and potentially intervening on,
587 multiple tasks, rather than relying on a single task to capture complex cognitive domains such
588 as g_f . Finally, the age-related differences in the relationships of the watershed model
589 observed using SEM-trees suggest that some interventions may work best at particular
590 developmental phases.

591 **Acknowledgements**

592 D.F. and J.B. are supported by the UK Medical Research Council (MRC). I.L.S.-K. is supported
593 by the Cambridge Trust. R.A.K. is supported by the Sir Henry Wellcome Trust (107392/Z/15/Z)
594 and MRC Programme Grant (MC-A060-5PR60). The Centre for Attention Learning and
595 Memory (CALM) research clinic is based at and supported by funding from the MRC

596 Cognition and Brain Sciences Unit, University of Cambridge. The Principal Investigators are
597 Joni Holmes (Head of CALM), Susan Gathercole (Chair of CALM Management Committee),
598 Duncan Astle, Tom Manly and Rogier Kievit. Data collection is assisted by a team of
599 researchers and PhD students at the CBU that includes Sarah Bishop, Annie Bryant, Sally
600 Butterfield, Fanchea Daily, Laura Forde, Erin Hawkins, Sinead O’Brien, Clodhna O’Leary,
601 Joseph Rennie, and Mengya Zhang. The authors wish to thank the many professionals
602 working in children’s services in the South-East and East of England for their support, and to
603 the children and their families for giving up their time to visit the clinic. We would also like to
604 thank all NKI-RS participants and researchers. We are grateful to Amber Ruigrok for helpful
605 suggestions regarding pubertal development.

606 **References**

- 607 Ackerman PL, Beier ME, Boyle MO. 2005. Working memory and intelligence: The same or
608 different constructs? *Psychol Bull.* 131:30–60.
- 609 Alloway TP. 2007. *Automated Working Memory Assessment*. London, UK: The Psychological
610 Corporation.
- 611 Bathelt J, Johnson A, Zhang M, Astle DE. 2018. Data-driven brain-types and their cognitive
612 consequences. *bioRxiv*.
- 613 Blum D, Holling H. 2017. Spearman’s law of diminishing returns. A meta-analysis. *Intelligence.*
614 65:60–66.
- 615 Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T,
616 Mehta P, Fox J. 2011. OpenMx: An open source extended Structural Equation
617 Modeling framework. *Psychometrika.* 76:306–317.

618 Brandmaier AM, von Oertzen T, McArdle JJ, Lindenberger U. 2013. Structural Equation Model
619 Trees. *Psychol Methods*. 18:71–86.

620 Byrne ML, Whittle S, Vijayakumar N, Dennison M, Simmons JG, Allen NB. 2017. A systematic
621 review of adrenarche as a sensitive period in neurobiological development and
622 mental health. *Sensitive Periods Dev*. 25:12–28.

623 Cannon TD, Keller MC. 2006. Endophenotypes in the genetic analyses of mental disorders.
624 *Annu Rev Clin Psychol*. 2:267–290.

625 Chevalier N, Kurth S, Doucette MR, Wiseheart M, Deoni SCL, Dean DC III, O’Muircheartaigh J,
626 Blackwell KA, Munakata Y, LeBourgeois MK. 2015. Myelination is associated with
627 processing speed in early childhood: Preliminary insights. *PLOS ONE*. 10:e0139897.

628 Chuderski A. 2013. When are fluid intelligence and working memory isomorphic and when
629 are they not? *Intelligence*. 41:244–262.

630 Cox SR, Ritchie SJ, Tucker-Drob EM, Liewald DC, Hagenaars SP, Davies G, Wardlaw JM, Gale
631 CR, Bastin ME, Deary IJ. 2016. Ageing and brain white matter structure in 3,513 UK
632 Biobank participants. *Nat Commun*. 7:13629.

633 Coyle TR, Pillow DR, Snyder AC, Kochunov P. 2011. Processing speed mediates the
634 development of general intelligence (g) in adolescence. *Psychol Sci*. 22:1265–1269.

635 de Mooij SMM, Henson RNA, Waldorp LJ, Kievit RA. 2018. Age differentiation within grey
636 matter, white matter and between memory and white matter in an adult lifespan
637 cohort. *J Neurosci*.

638 Delis DC, Kramer JH, Kaplan E, Holdnack J. 2004. Reliability and validity of the Delis-Kaplan
639 Executive Function System: An update. *J Int Neuropsychol Soc*. 10:301–303.

640 Dorn LD. 2006. Measuring puberty. *J Adolesc Health*. 39:625–626.

641 Dunning DL, Holmes J, Gathercole SE. 2013. Does working memory training lead to
642 generalized improvements in children with low working memory? A randomized
643 controlled trial. *Dev Sci.* 16:915–925.

644 Fedorenko E, Duncan J, Kanwisher N. 2013. Broad domain generality in focal regions of
645 frontal and parietal cortex. *Proc Natl Acad Sci.* 201315235.

646 Ferrer E, Whitaker K, Steele J, Green C, Wendelken C, Bunge S. 2013. White matter
647 maturation supports the development of reasoning ability through its influence on
648 processing speed. *Dev Sci.* 16:941–951.

649 Finkel D, Reynolds CA, McArdle JJ, Pedersen NL. 2005. The longitudinal relationship between
650 processing speed and cognitive ability: Genetic and environmental influences. *Behav*
651 *Genet.* 35:535–549.

652 Fry AF, Hale S. 2000. Relationships among processing speed, working memory, and fluid
653 intelligence in children. *Biol Psychol.* 54:1–34.

654 Fukuda K, Vogel E, Mayr U, Awh E. 2010. Quantity not quality: The relationship between fluid
655 intelligence and working memory capacity. *Psychon Bull Rev.* 17:673–679.

656 Gale CR, Batty GD, Tynelius P, Deary IJ, Rasmussen F. 2010. Intelligence in early adulthood
657 and subsequent hospitalisation and admission rates for the whole range of mental
658 disorders: Longitudinal study of 1,049,663 men. *Epidemiol Camb Mass.* 21:70–77.

659 Gallagher A, Frederickson N. 1995. The Phonological Assessment Battery (PhAB): An initial
660 assessment of its theoretical and practical utility. *Educ Child Psychol.*

661 Garrett HE. 1946. A developmental theory of intelligence. *Am Psychol.* 1:372–378.

662 Gignac GE, Szodorai ET. 2016. Effect size guidelines for individual differences researchers.
663 *Personal Individ Differ.* 102:74–78.

664 Gur RC, Ragland JD, Moberg PJ, Turner TH, Bilker WB, Kohler C, Siegel SJ, Gur RE. 2001.
665 Computerized Neurocognitive Scanning: I. methodology and validation in healthy
666 people. *Neuropsychopharmacology*. 25:766.

667 Holmes J, Bryant A, Gathercole SE. 2018. A transdiagnostic study of children with problems of
668 attention, learning and memory (CALM). *bioRxiv*.

669 Hua K, Zhang J, Wakana S, Jiang H, Li X, Reich DS, Calabresi PA, Pekar JJ, van Zijl PCM, Mori S.
670 2008. Tract probability maps in stereotaxic spaces: analyses of white matter anatomy
671 and tract-specific quantification. *NeuroImage*. 39:336–347.

672 Johnson MH. 2000. Functional brain development in infants: Elements of an interactive
673 specialization framework. *Child Dev*. 71:75–81.

674 Johnson MH. 2011. Interactive Specialization: A domain-general framework for human
675 functional brain development? *Dev Cogn Neurosci*. 1:7–21.

676 Jones DK, Knösche TR, Turner R. 2013. White matter integrity, fiber count, and other fallacies:
677 The do's and don'ts of diffusion MRI. *NeuroImage*. 73:239–254.

678 Jöreskog KG. 1999. How large can a standardized coefficient be?

679 Jung RE, Haier RJ. 2007. The Parieto-Frontal Integration Theory (P-FIT) of intelligence:
680 Converging neuroimaging evidence. *Behav Brain Sci*. 30:135–154.

681 Kail RV, Lervåg A, Hulme C. 2015. Longitudinal evidence linking processing speed to the
682 development of reasoning. *Dev Sci*. 19:1067–1074.

683 Kail RV, Salthouse TA. 1994. Processing speed as a mental capacity. *Acta Psychol (Amst)*.
684 86:199–225.

685 Kamali A, Flanders AE, Brody J, Hunter JV, Hasan KM. 2014. Tracing superior longitudinal
686 fasciculus connectivity in the human brain using high resolution diffusion tensor
687 tractography. *Brain Struct Funct*. 219:10.1007/s00429-012-0498-y.

688 Kaufman AS. 1975. Factor analysis of the WISC-R at 11 age levels between 61/2 and 161/2
689 years. *J Consult Clin Psychol.* 43:135–147.

690 Kievit R, Fuhrmann D, Borgeest G, Simpson-Kent I, Henson R. 2018. The neural determinants
691 of age-related changes in fluid intelligence: a pre-registered, longitudinal analysis in
692 UK Biobank [version 2; referees: 3 approved]. *Wellcome Open Res.* 3.

693 Kievit RA, Davis SW, Griffiths J, Correia MM, CamCAN, Henson RN. 2016. A watershed model
694 of individual differences in fluid intelligence. *Neuropsychologia.* 91:186–198.

695 Kievit RA, Davis SW, Griffiths J, Correia MM, Cam-CAN, Henson RN. 2016. A watershed model
696 of individual differences in fluid intelligence. *Neuropsychologia.* 91:186–198.

697 Kievit RA, Romeijn JW, Waldorp LJ, Wicherts JM, Scholte HS, Borsboom D. 2011. Mind the
698 gap: A psychometric approach to the reduction problem. *Psychol Inq.* 22:67–87.

699 Kline RB. 2015. *Principles and practice of structural equation modeling.* Guilford publications.

700 Knoll LJ, Fuhrmann D, Sakhardande AL, Stamp F, Speekenbrink M, Blakemore SJ. 2016. A
701 window of opportunity for cognitive training in adolescence. *Psychol Sci.* 27:1620–
702 1631.

703 Kyllonen PC, Christal RE. 1990. Reasoning ability is (little more than) working-memory
704 capacity?! *Intelligence.* 14:389–433.

705 Lövdén M, Laukka EJ, Rieckmann A, Kalpouzos G, Li T, Jonsson T, Wahlund L, Fratiglioni L,
706 Bäckman L. 2013. The dimensionality of between-person differences in white matter
707 microstructure in old age. *Hum Brain Mapp.* 34:1386–1398.

708 Lu PH, Lee GJ, Raven EP, Tingus K, Khoo T, Thompson PM, Bartzokis G. 2011. Age-related
709 slowing in cognitive processing speed is associated with myelin integrity in a very
710 healthy elderly sample. *J Clin Exp Neuropsychol.* 33:1059–1068.

711 Mackey AP, Hill SS, Stone SI, Bunge SA. 2011. Differential effects of reasoning and speed
712 training in children. *Dev Sci.* 14:582–590.

713 MacPherson SE, Cox SR, Dickie DA, Karama S, Starr JM, Evans AC, Bastin ME, Wardlaw JM,
714 Deary IJ. 2017. Processing speed and the relationship between Trail Making Test-B
715 performance, cortical thinning and white matter microstructure in older adults.
716 *Cortex.* 95:92–103.

717 Manly T, Anderson V, Nimmo-Smith I, Turner A, Watson P, Robertson IH. 2001. The
718 differential assessment of children’s attention: The Test of Everyday Attention for
719 Children (TEA-Ch), normative sample and ADHD performance. *J Child Psychol*
720 *Psychiatry.* 42:1065–1081.

721 Mills KL, Goddings A-L, Herting MM, Meuwese R, Blakemore S-J, Crone EA, Dahl RE, Güroğlu
722 B, Raznahan A, Sowell ER, Tamnes CK. 2016. Structural brain development between
723 childhood and adulthood: Convergence across four longitudinal samples.
724 *NeuroImage.* 141:273–281.

725 Mills KL, Tamnes CK. 2014. Methods and considerations for longitudinal structural brain
726 imaging analysis across development. *Dev Cogn Neurosci.* 9:172–190.

727 Noack H, Lövdén M, Schmiedek F. 2014. On the validity and generality of transfer effects in
728 cognitive training research. *Psychol Res.* 78:773–789.

729 Nooner K, Colcombe S, Tobe R, Mennes M, Benedict M, Moreno A, Panek L, Brown S, Zavitz
730 S, Li Q, Sikka S, Gutman D, Bangaru S, Schlachter RT, Kamiel S, Anwar A, Hinz C, Kaplan
731 M, Rachlin A, Adelsberg S, Cheung B, Khanuja R, Yan C, Craddock C, Calhoun V,
732 Courtney W, King M, Wood D, Cox C, Kelly C, DiMartino A, Petkova E, Reiss P, Duan N,
733 Thompsen D, Biswal B, Coffey B, Hoptman M, Javitt D, Pomara N, Sidtis J, Koplewicz H,

734 Castellanos F, Leventhal B, Milham M. 2012. The NKI-Rockland Sample: A model for
735 accelerating the pace of discovery science in psychiatry. *Front Neurosci.* 6:152.

736 Oberauer K, Schulze R, Wilhelm O, Süß H-M. 2005. Working memory and intelligence - Their
737 correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005).
738 *Psychol Bull.* 131:61–65.

739 Orben A, Przybylski AK. 2019. The association between adolescent well-being and digital
740 technology use. *Nat Hum Behav.*

741 Parlatini V, Radua J, Dell’Acqua F, Leslie A, Simmons A, Murphy DG, Catani M, Thiebaut de
742 Schotten M. 2017. Functional segregation and integration within fronto-parietal
743 networks. *NeuroImage.* 146:367–375.

744 Penke L, Maniega SM, Murray C, Gow AJ, Valdés Hernández MC, Clayden JD, Starr JM,
745 Wardlaw JM, Bastin ME, Deary IJ. 2010. A general factor of brain white matter
746 integrity predicts information processing speed in healthy older people. *J Neurosci.*
747 30:7569.

748 Pfefferbaum A, Sullivan EV, Hedehus M, Lim KO, Adalsteinsson E, Moseley M. 2000. Age-
749 related decline in brain white matter anisotropy measured with spatially corrected
750 echo-planar diffusion tensor imaging. *Magn Reson Med.* 44:259–268.

751 Primi R, Ferrão ME, Almeida LS. 2010. Fluid intelligence as a predictor of learning: A
752 longitudinal multilevel approach applied to math. *Learn Individ Differ.* 20:446–451.

753 R core team the. 2015. R: A language and environment for statistical computing.

754 Ritchie SJ, Tucker-Drob EM. 2018. How much does education improve intelligence? A meta-
755 analysis. *Psychol Sci.* 0956797618774253.

756 Rosseel Y. 2012. lavaan: An R package for Structural Equation Modeling. *J Stat Softw.* 48:1–
757 36.

758 Roth B, Becker N, Romeyke S, Schäfer S, Domnick F, Spinath FM. 2015. Intelligence and
759 school grades: A meta-analysis. *Intelligence*. 53:118–137.

760 Salthouse TA. 1996. The processing-speed theory of adult age differences in cognition.
761 *Psychol Rev*. 103:403–428.

762 Schermelleh-Engel K, Moosbrugger H, Müller H. 2003. Evaluating the fit of Structural
763 Equation Models: Tests of significance and descriptive goodness-of-fit measures.
764 *Methods Psychol Res*. 8:23–74.

765 Schubert A-L, Hagemann D, Frischkorn GT. 2017. Is general intelligence little more than the
766 speed of higher-order processing? *J Exp Psychol Gen*. 146:1498–1512.

767 Schwaighofer M, Fischer F, Bühner M. 2015. Does working memory training transfer? A
768 meta-analysis including training conditions as moderators. *Educ Psychol*. 50:138–166.

769 Scott KM, Kline M. 2019. Enabling Confirmatory Secondary Data Analysis by Logging Data
770 Checkout. *Adv Methods Pract Psychol Sci*. 2515245918815849.

771 Sheppard LD, Vernon PA. 2008. Intelligence and speed of information-processing: A review of
772 50 years of research. *Personal Individ Differ*. 44:535–551.

773 Simpson-Kent IL, Fuhrmann D, Bathelt J, the CALM team, Kievit RA. 2019. Cross-sectional
774 evidence for age differentiation-dedifferentiation in crystallized and fluid intelligence
775 and their white matter substrates during childhood and adolescence. *Prep*.

776 Sisk CL, Zehr JL. 2005. Pubertal hormones organize the adolescent brain and behavior. *Front*
777 *Neuroendocrinol*. 26:163–174.

778 Tamnes CK, Roalf DR, Goddings A-L, Lebel C. 2017. Diffusion MRI of white matter
779 microstructure development in childhood and adolescence: Methods, challenges and
780 progress. *Dev Cogn Neurosci*.

781 Tucker-Drob EM. 2009. Differentiation of cognitive abilities across the lifespan. *Dev Psychol.*
782 45:1097–1118.

783 Vestergaard M, Madsen KS, Baaré WF, Skimminge A, Ejersbo LR, Ramsøy TZ, Gerlach C,
784 Åkeson P, Paulson OB, Jernigan TL. 2011. White matter microstructure in superior
785 longitudinal fasciculus associated with spatial working memory performance in
786 children. *J Cogn Neurosci.* 23:2135–2146.

787 Volkow ND, Koob GF, Croyle RT, Bianchi DW, Gordon JA, Koroshetz WJ, Pérez-Stable EJ, Riley
788 WT, Bloch MH, Conway K, Deeds BG, Dowling GJ, Grant S, Howlett KD, Matochik JA,
789 Morgan GD, Murray MM, Noronha A, Spong CY, Wargo EM, Warren KR, Weiss SRB.
790 2018. The conception of the ABCD study: From substance use to a broad NIH
791 collaboration. *Adolesc Brain Cogn Dev ABCD Consort Ration Aims Assess Strategy.*
792 32:4–7.

793 Vollmar C, O’Muircheartaigh J, Barker GJ, Symms MR, Thompson P, Kumari V, Duncan JS,
794 Richardson MP, Koepp MJ. 2010. Identical, but not the same: intra-site and inter-site
795 reproducibility of fractional anisotropy measures on two 3.0T scanners. *NeuroImage.*
796 51:1384–1394.

797 Wagenmakers E-J, Farrell S. 2004. AIC model selection using Akaike weights. *Psychon Bull*
798 *Rev.* 11:192–196.

799 Wagenmakers E-J, Wetzels R, Borsboom D, van der Maas HLJ, Kievit RA. 2012. An agenda for
800 purely confirmatory research. *Perspect Psychol Sci.* 7:632–638.

801 Wandell BA. 2016. Clarifying human white matter. *Annu Rev Neurosci.* 39:103–128.

802 Wechsler D. 2011. *Wechsler Abbreviated Scale of Intelligence -Second Edition.* San Antonio,
803 TX: Pearson.

804 Weston SJ, Ritchie SJ, Rohrer JM, Przybylski AK. 2018. Recommendations for increasing the
805 transparency of analysis of pre-existing datasets. PsyArXiv.

806

807

Figure Legends

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

813 *Figure 2. White Matter Tracts Modelled in the Analyses.*

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

816 *Figure 4. The Watershed Model in CALM.* See Supplementary Table 2 for regression
817 estimates. Residual covariances between white matter tracts were allowed but are not
818 shown for simplicity. Abbreviations: uncinat fasciculus (UF), superior longitudinal fasciculus
819 (SLF), inferior fronto-occipital fasciculus (IFOF), anterior thalamic radiations (ATR),
820 cerebrosipinal tract (CST), forceps major (FMaj), forceps minor (FMin), dorsal cingulate gyrus
821 (CG), ventral cingulate gyrus (CH), inferior longitudinal fasciculus (ILF).

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

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

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