

1 Neurocognitive reorganization between crystallized intelligence, 2 fluid intelligence and white matter microstructure in two age- 3 heterogeneous developmental cohorts

4 Ivan L. Simpson-Kent ^{a, *}, Delia Fuhrmann ^a, Joe Bathelt ^b, Jascha Achterberg ^{a, †}, Gesa Sophia
5 Borgeest ^{a, †}, the CALM Team and Rogier A. Kievit ^a

6 ^a MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, Cambridgeshire,
7 CB2 7EF, UK

8 ^b Dutch Autism & ADHD Research Center, Brain & Cognition, University of Amsterdam, 1018 WS
9 Amsterdam, Netherlands

10 * Correspondence: Ivan.Simpson-Kent@mrc-cbu.cam.ac.uk; 15 Chaucer Road, Cambridge CB2
11 7EF, UK; Tel.: +44 (0) 1223 769899

12 † Equal contributions

13 14 Abstract

15
16 Despite the reliability of intelligence measures in predicting important life outcomes such as
17 educational achievement and mortality, the exact configuration and neural correlates of
18 cognitive abilities remain poorly understood, especially in childhood and adolescence.
19 Therefore, we sought to elucidate the factorial structure and neural substrates of child and
20 adolescent intelligence using two cross-sectional, developmental samples (CALM: N=551,
21 age range: 5-18 years, NKI: N=337, age range: 6-18 years). In a preregistered analysis, we
22 used structural equation modelling (SEM) to examine the neurocognitive architecture of
23 individual differences in childhood and adolescent cognitive ability. In both samples, we
24 found that cognitive ability in lower and typical-ability cohorts is best understood as two
25 separable constructs, crystallized and fluid intelligence, which became more distinct across
26 development. Further analyses revealed that white matter microstructure, most prominently
27 the superior longitudinal fasciculus, was strongly associated with crystallized (gc) and fluid
28 (gf) abilities. Finally, we used SEM trees to demonstrate evidence for age differentiation-
29 dedifferentiation of gc and gf and their white matter substrates such that the relationships
30 among these factors dropped between 7-8 years before increasing around age 10. Together,
31 our results suggest that shortly before puberty marks a pivotal phase of change in the
32 neurocognitive architecture of intelligence.

33 34 Keywords

35 Age differentiation-dedifferentiation; Crystallized intelligence; Fluid intelligence; White matter;
36 Structural equation modelling

37
38

39 1. Introduction

40

41 Intelligence measures have repeatedly been shown to predict important life
42 outcomes such as educational achievement (Deary et al., 2007) and mortality (Calvin et al.,
43 2011). Modern investigations of intelligence began over 100 years ago, when Spearman first
44 proposed *g* (for 'general intelligence') as the underlying factor behind his positive manifold of
45 cognitive ability and established intelligence as a central theme of psychological research
46 (Spearman, 1904). Cattell proposed a division of Spearman's *g*-factor into two separate yet
47 related constructs, crystallized (*gc*) and fluid (*gf*) intelligence (Cattell, 1967). Cattell
48 suggested that *gc* represents the capacity to effectively complete tasks based on acquired
49 knowledge and experience (e.g. arithmetic, vocabulary) whereas *gf* refers to one's ability to
50 solve novel problems without task-specific knowledge, relying on abstract thinking and
51 pattern recognition (see also Deary et al., 2010).

52 Current understanding of lifespan trajectories of *gc* and *gf* using cross-sectional
53 (Horn and Cattell, 1967) and longitudinal (McArdle et al., 2000; Schaie, 1994) cohorts
54 indicates that *gc* slowly improves until late age while *gf* increases into early adulthood before
55 steadily decreasing. However, the majority of the literature on individual differences between
56 *gc* and *gf* has focused on early to late adulthood. As a result, considerably less is known
57 about the association between *gc* and *gf* in childhood and adolescence (but see Hülür et al.,
58 2011).

59 There has, however, been a recent rise in interest in this topic in child and adolescent
60 samples. For instance, research on age-related differentiation and its inverse, age
61 dedifferentiation, in younger samples has greatly expanded since first being pioneered in the
62 middle of the 20th century (Garrett, 1946). According to the age differentiation hypothesis,
63 cognitive factors become less correlated (more differentiated) with increasing age. For
64 example, the relationship (covariance) between *gc* and *gf* would decrease as children age
65 into adolescence, suggesting that cognitive abilities increasingly specialize into adulthood. In
66 contrast, the age dedifferentiation hypothesis predicts that cognitive abilities become more
67 strongly related (less differentiated) throughout development. In this case, *gc* and *gf*
68 covariance would increase between childhood and adolescence, potentially indicating a
69 strengthening of the *g*-factor across age. However, despite its increased attention in the
70 literature, the debate remains unsolved as evidence in support of both hypotheses has been
71 found (Bickley et al., 1995; de Mooij et al., 2018; Gignac, 2014; Hülür et al., 2011; Juan-
72 Espinosa et al., 2000; Tideman and Gustafsson, 2004). Together, this literature highlights
73 the importance of a lifespan perspective on theories of cognitive development, as neither
74 age differentiation nor dedifferentiation may be solely able to capture the dynamic changes
75 that occur from childhood to adolescence and (late) adulthood (Hartung et al., 2018).

76 The introduction of non-invasive brain imaging technology has complemented
77 conventional psychometric approaches by allowing for fine-grained probing of the neural
78 bases of human cognition. A particular focus in developmental cognitive neuroscience has
79 been the study of white matter using techniques such as diffusion-weighted imaging, which
80 allows for the estimation of white matter microstructure (Wandell, 2016). Both cross-
81 sectional and longitudinal research in children and adolescents using fractional anisotropy
82 (FA), a commonly used estimate of white matter integrity, have consistently revealed positive
83 correlations between FA and cognitive ability using tests of working memory, verbal and
84 non-verbal performance (Krogsrud et al., 2018; Peters et al., 2014; Tamnes et al., 2010;
85 Urger et al., 2015). In particular, recent research has found associations between the corpus
86 callosum (Navas-Sánchez et al., 2014; Westerhausen et al., 2018) association fibers (e.g.
87 inferior longitudinal fasciculus, see Peters et al., 2014), the superior longitudinal fasciculus
88 (Urger et al., 2015), and differences in cognitive ability, suggesting the importance of white
89 matter integrity across large coordinated brain networks for high cognitive performance.
90 However, interpretations of these studies are limited due to restricted cognitive batteries
91 (e.g. small number of tests used) and a dearth of theory-driven statistical analyses (e.g.
92 structural equation modelling).

93 For these reasons, several outstanding questions in the developmental cognitive
94 neuroscience of intelligence remain: 1) Are the white matter substrates underlying
95 intelligence in childhood and adolescence best understood as a single global factor or do
96 individual tracts provide specific contributions to gc and gf?, 2) If they are specific, are the
97 tract contributions identical between gc and gf?, and 3) Does this brain-behavior mapping
98 change in development (e.g. age differentiation/dedifferentiation or both)?

99 To examine these questions, our [preregistered](#) hypotheses are as follows:

- 100 1) gc and gf are separable constructs in childhood and adolescence. More
101 specifically, the covariance among scores on cognitive tests are more adequately
102 captured by the two-factor (gc-gf) model as opposed to a single-factor (e.g. g)
103 model.
- 104 2) The covariance between gc and gf changes (decreases) across childhood and
105 adolescence.
- 106 3) White matter tracts make unique complementary contributions to gc and gf.
- 107 4) The contributions of these tracts to gc and gf change (decrease) with age.

108 To address these questions, we examined the relationship between gc and gf in two
109 large cross-sectional child and adolescent samples. The first is the Centre for Attention,
110 Learning and Memory (CALM, see Holmes et al., 2019). This sample, included in our
111 [preregistration](#), was recruited atypically (see Methods for more detail) and generally includes
112 children with slightly lower cognitive abilities than age-matched controls. To examine

113 whether findings from CALM would generalize to other samples, we also conducted non-
114 preregistered analyses on the Nathan Kline Institute (NKI) Rockland Sample, a cohort with
115 similar population demographics to the United States (e.g. race and socioeconomic status,
116 see Table 1 of Nooner et al., 2012). All analyses were carried out using structural equation
117 modelling (SEM), a multivariate statistical framework combining factor and path analysis to
118 examine the extent to which causal hypotheses concerning latent (unobserved, e.g. *g*) and
119 manifest (observed, e.g. cognitive tests scores) variables (Schreiber et al., 2006) are in line
120 with the observed data. Taken together, this paper sought to investigate the relationship
121 between measures of intelligence (*gc* and *gf*) and white matter connectivity in typically and
122 atypically (struggling learners) developing children and adolescents.

123

124 **2. Methods**

125

126 *2.1.1 Participants*

127

128 For the CALM sample, we analyzed the most recent data release (N=551; 170
129 female, 381 male¹; age range=5.17-17.92 years) at the time of preregistration (see
130 <https://aspredicted.org/5pz52.pdf>). Participants were recruited based on referrals made for
131 possible attention, memory, language, reading and/or mathematics problems (Holmes et al.,
132 2019). Participants with or without formal clinical diagnosis were referred to CALM.
133 Exclusion criteria included known significant and uncorrected problems in vision or hearing
134 and a native language other than English. A subset of participants completed MRI scanning
135 (N=165; 56 female, 109 male; age range=5.92-17.92 years). For more information about
136 CALM, see <http://calm.mrc-cbu.cam.ac.uk/>.

137 Next, to assess the generalizability of our findings in CALM, we used a non-
138 preregistered subset of the data from the Nathan Kline Institute (NKI) Rockland Sample
139 (cognitive data: N=337; 149 female, 188 male; age range=6.12-17.94 years; neural data:
140 N=65; 27 female, 38 male; age range=6.97-17.8 years). This multi-institutional initiative
141 recruited a lifespan (aged between 6 and 85 years), community-ascertained sample (Nooner
142 et al., 2012). We chose this sample due to its representativeness (demographics resemble
143 those of the United States population) and the fact that its cognitive battery assessments
144 closely-matched CALM. For more information about the NKI Rockland Sample and its
145 procedures, see <http://rocklandsample.org/>. Also see Fig. 1 for age distributions of CALM

¹ Gender was coded as either female or male. However, it should be noted that participants might identify themselves as 'Other', which, to our knowledge, was not an option according to the biographical produces used in either sample.

146 and NKI. These same two cohorts were used in a recent paper to address a distinct set of
147 questions (Fuhrmann et al., 2019).

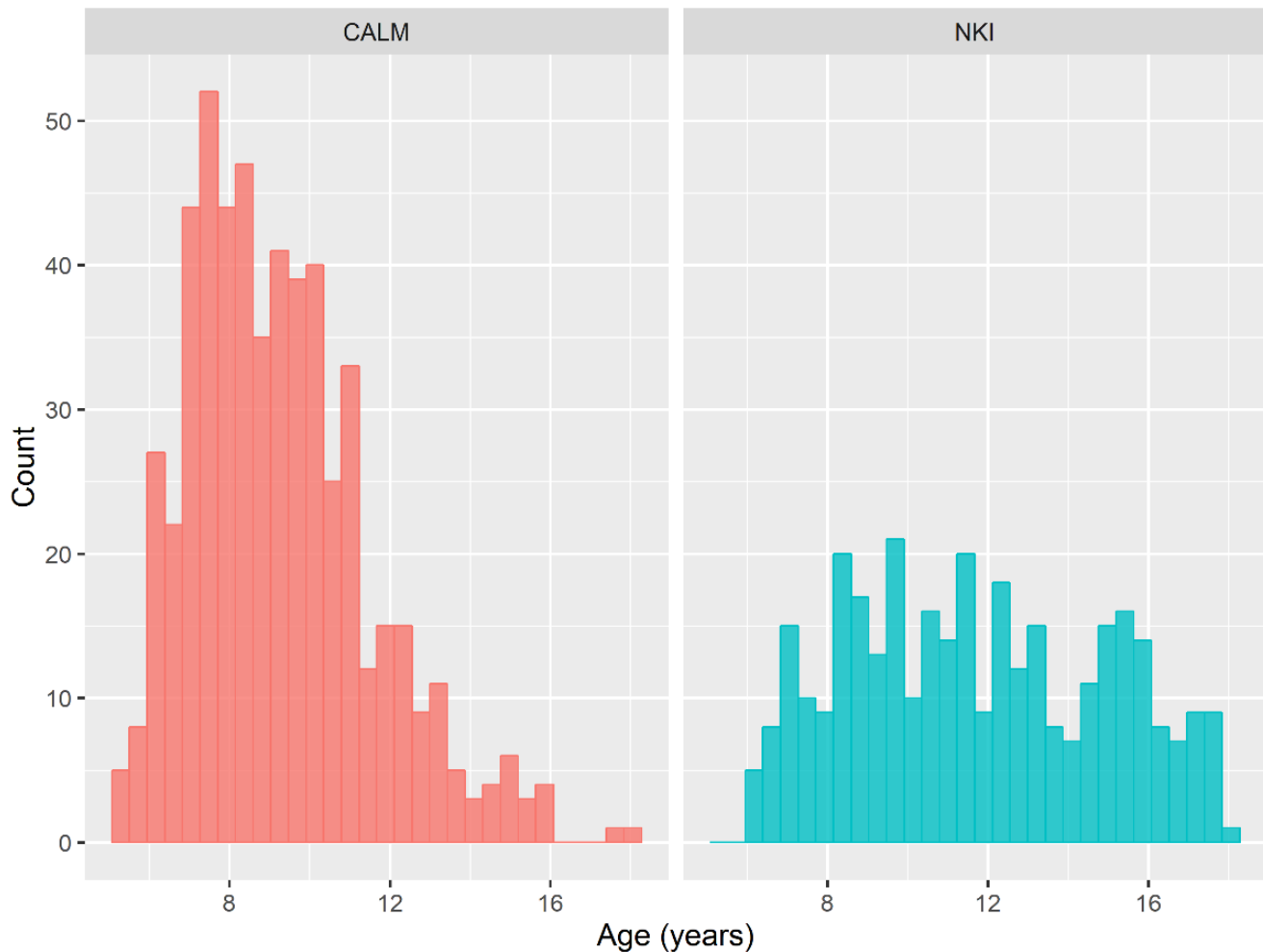


Fig. 1. Histograms of age distributions for CALM and NKI Rockland samples.

148 2.1.2 Statistical analyses

149

150 We used structural equation modelling (SEM), a multivariate approach that combines
151 latent variables and path modelling to test causal hypotheses (Schreiber et al., 2006) as well
152 as SEM trees, which combine SEM and decision tree paradigms to simultaneously permit
153 exploratory and confirmatory data analysis (Brandmaier et al., 2013).

154 We performed structural equation modelling (SEM) using the lavaan package version
155 0.5-22 (Rosseel, 2012) in R (R Core Team, 2018) and versions 2.9.9 and 0.9.12 of the R
156 packages OpenMx (Boker et al., 2011) and semtree (Brandmaier et al., 2013), respectively.
157 To account for missing data and deviations from multivariate normality, we used robust full
158 information maximum likelihood estimator (FIML) with a Yuan-Bentler scaled test statistic
159 (MLR) and robust standard errors (Rosseel, 2012). We evaluated overall model fit via the

160 (Satorra-Bentler scaled) chi-squared test, the comparative fit index (CFI), the standardized
161 root mean squared residuals (SRMR), and the root mean square error of approximation
162 (RMSEA) with its confidence interval (Schermele-Engel et al., 2003). Assessment of model
163 fit was defined as: CFI (acceptable fit 0.95-0.97, good fit >0.97), SRMR (acceptable fit 0.05-
164 .10, good fit <0.05), and RMSEA (acceptable fit 0.05-0.08, good fit <0.05). To determine
165 whether gc and gf were separable constructs, we compared a two-factor (gc-gf) model to an
166 single-factor (g) model. To investigate if the covariance between gc and gf differed across
167 ages, we conducted multiple group comparisons between younger and older participants
168 based on median splits (CALM split at 8.91 years yielding N=279 young and 272 old; NKI
169 split at 11.38 years into N=169 young and N=168 old). Doing so inevitably led to slightly
170 unbalanced numbers of participants with white matter data (CALM: young, N=60 & old,
171 N=105; NKI: young, N=19 & old, N=46). To test measurement invariance across age groups
172 (Putnick and Bornstein, 2016), we fit multigroup models (French and Finch, 2008),
173 constraining key parameters across groups. Model comparisons and deviations from
174 measurement invariance were determined using the likelihood ratio test and Akaike
175 information criterion (AIC, see Bozdogan, 1987).

176 To examine whether white matter tracts made unique contributions to our latent
177 variables we fit Multiple Indicator, Multiple Cause (MIMIC) models (Jöreskog and
178 Goldberger, 1975; Kievit et al., 2012). Lastly, we conducted a SEM tree analysis, a method
179 that combines the confirmatory nature of SEM with the exploratory framework of decision
180 trees (Brandmaier et al., 2013). SEM trees hierarchically and recursively partition data
181 (decision tree) according to covariates that explain the maximum difference in parameter
182 estimates of a theorized model (SEM). For each SEM tree analysis, a minimum sample size
183 of 100 was set for each node to aid estimation. Our use of this technique was twofold: 1)
184 Examine the robustness of findings based on the median age split, and 2) examine whether
185 white matter contributions differed across age groups of younger and older participants
186 (Hypothesis 4). Therefore, for our SEM tree analyses in CALM and NKI, we used age as a
187 continuous covariate.

188

189 *2.1.3 Cognitive assessments: gc, gf, and working memory*

190

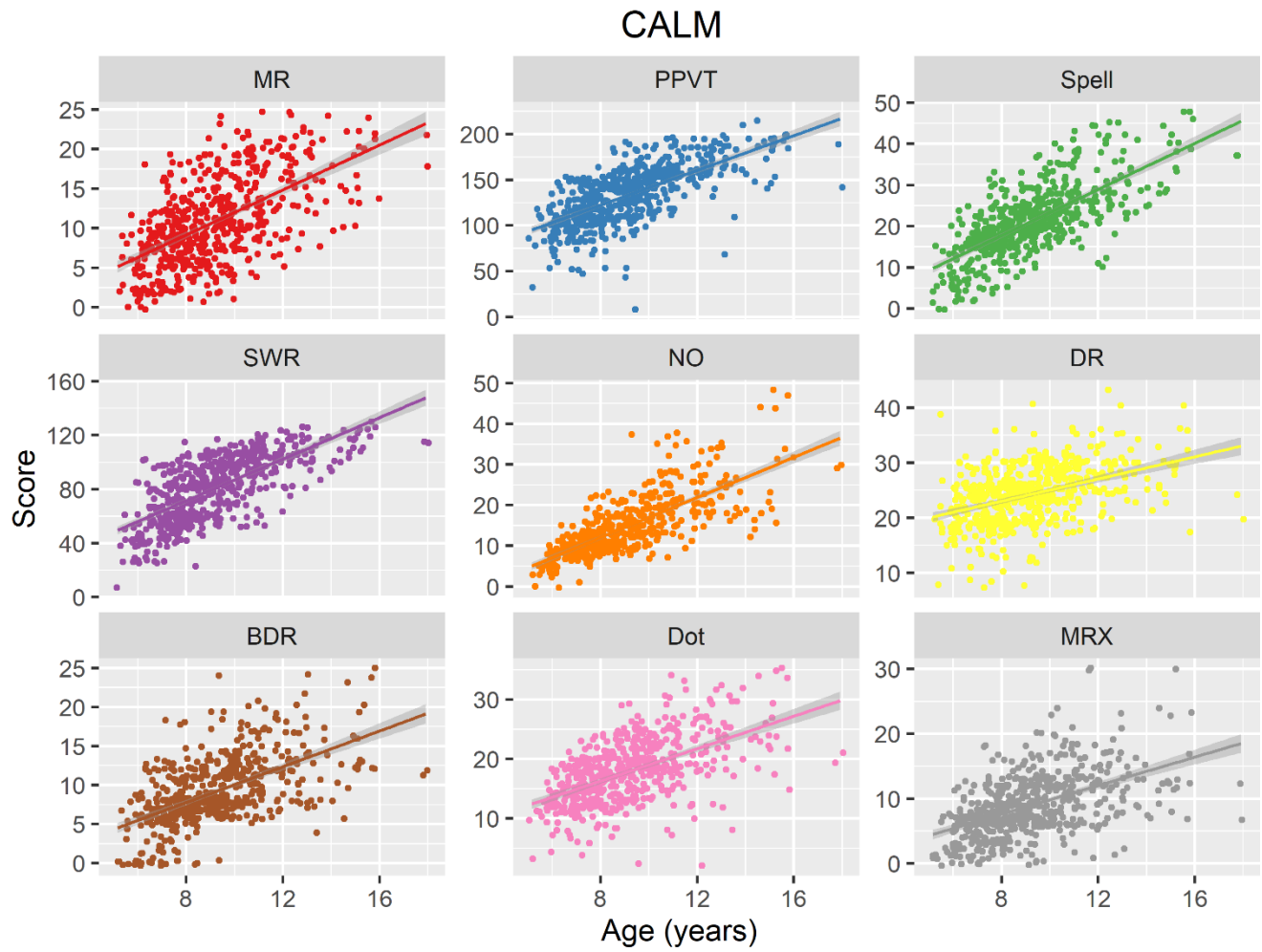
191 All cognitive data from the CALM sample were collected on a one-to-one basis by an
192 examiner in a dedicated child-friendly testing room. The test battery included a wide range of
193 standardized assessments of cognition and learning (Holmes et al., 2019). Participants were
194 given regular breaks throughout the session. Testing was divided into two sessions for
195 participants who struggled to complete the assessments in one sitting. For analyses of the
196 NKI Rockland Sample cohort, we matched tasks used in CALM except for the Peabody

197 Picture Vocabulary Test, Dot Matrix, and Mr. X, which were only available for CALM. For the
 198 NKI Rockland Sample, we included the N-Back task, which is not available in CALM (Nooner
 199 et al., 2012). In both samples, only raw scores obtained from assessments were included in
 200 analyses. Due to varying delays between recruitment and testing in NKI, we only used
 201 cognitive test scores completed no later than six months after initial recruitment. The
 202 cognitive tasks are further described in Table 1; the raw scores are depicted in Fig. 2.
 203

Cognitive Domain	Sample	Task and Description	Mean (sd) [range]	Missing Data %	Reference
Crystallized Ability (gc)	CALM	Peabody Picture Vocabulary Test (PPVT): Participants were asked to choose the picture (out of four multiple-choice options) showing the meaning of a word spoken by an examiner.	CALM: 133.77 (31.68) [8, 215] NKI: N/A	CALM: 1.09 NKI: N/A	Dunn and Dunn, 2007
	CALM & NKI	Single Word Reading (SWR): Participants read aloud first a list of letters and then words that gradually increased in complexity. Correct responses required correctness and fluency.	CALM: 80.95 (24.35) [7, 130] NKI: 104.47 (20.28) [35, 131]	CALM: 2.36 NKI: 0	Wechsler, 2005
	CALM & NKI	Spelling (Spell): Participants spelled words with increasing difficulty one at a time that were spoken by an examiner.	CALM: 21.17 (8.68) [0, 48] NKI: 33.57 (10.55) [4, 52]	CALM: 3.09 NKI: 0	
	CALM & NKI	Numerical Operations (NO): Participants answered written mathematical problems that increased in difficulty.	CALM: 14.83 (7.46) [0, 48] NKI: 27.95 (11.95) [4, 53]	CALM: 13.61 NKI: 0	
Fluid Ability (gf)	CALM & NKI	Matrix Reasoning (MR): Participants saw sequences of partial matrices and selected the response option that best completed each matrix.	CALM: 10.88 (5.44) [0, 25] NKI: 17.37 (5.19) [4, 27]	CALM: 0 NKI: 0	Wechsler, 1999 Wechsler, 2011
Working Memory (WM)	CALM & NKI	Digit Recall/Span (DR): Participants recalled sequences of single digit numbers given in audio format.	CALM: 24.22 (5.32) [7, 43] NKI: 5.97 (1.25) [3, 9]	CALM: 0.36 NKI: 24.63	Alloway, 2007 Kaufman, 1975

CALM & NKI	Backward Digit Recall/Span (BDR): Same as regular digit recall/span but in reversed order.	CALM: 9.2 (4.42) [0, 25] NKI: 4.04 (1.40) [0, 8]	CALM: 1.63 NKI: 24.63	
CALM	Dot Matrix (Dot): For 2 seconds, participants were shown the location of a red dot in a sequence of 4x4 matrices and had to recollect this location by tapping the squares on a computer screen.	CALM: 17.94 (5.49) [2, 35] NKI: N/A	CALM: 0.18 NKI: N/A	
CALM	Mr. X (MRX): Participants remembered spatial locations of a ball held by a cartoon man rotated in one of seven positions.	CALM: 8.94 (4.90) [0, 30] NKI: N/A	CALM: 0.91 NKI: N/A	
NKI	N-Back (NB): For 500 ms participants were presented letter sequences with a further 2000 ms to respond by pressing the computer spacebar. The task consisted of three separate conditions: 0-Back– participants pressed the spacebar whenever an “X” appeared; 1-Back– participants pressed the spacebar whenever the same letter was presented twice in a row; and, lastly, 2-Back– participants pressed the spacebar each time the letter presented matched the one shown two letters beforehand.	CALM: N/A NKI: 16.32 (4.22) [0, 20]	CALM: N/A NKI: 20.47	Gur et al., 2010

Table 1. List and Descriptions of Cognitive Assessments used in CALM & NKI Rockland samples



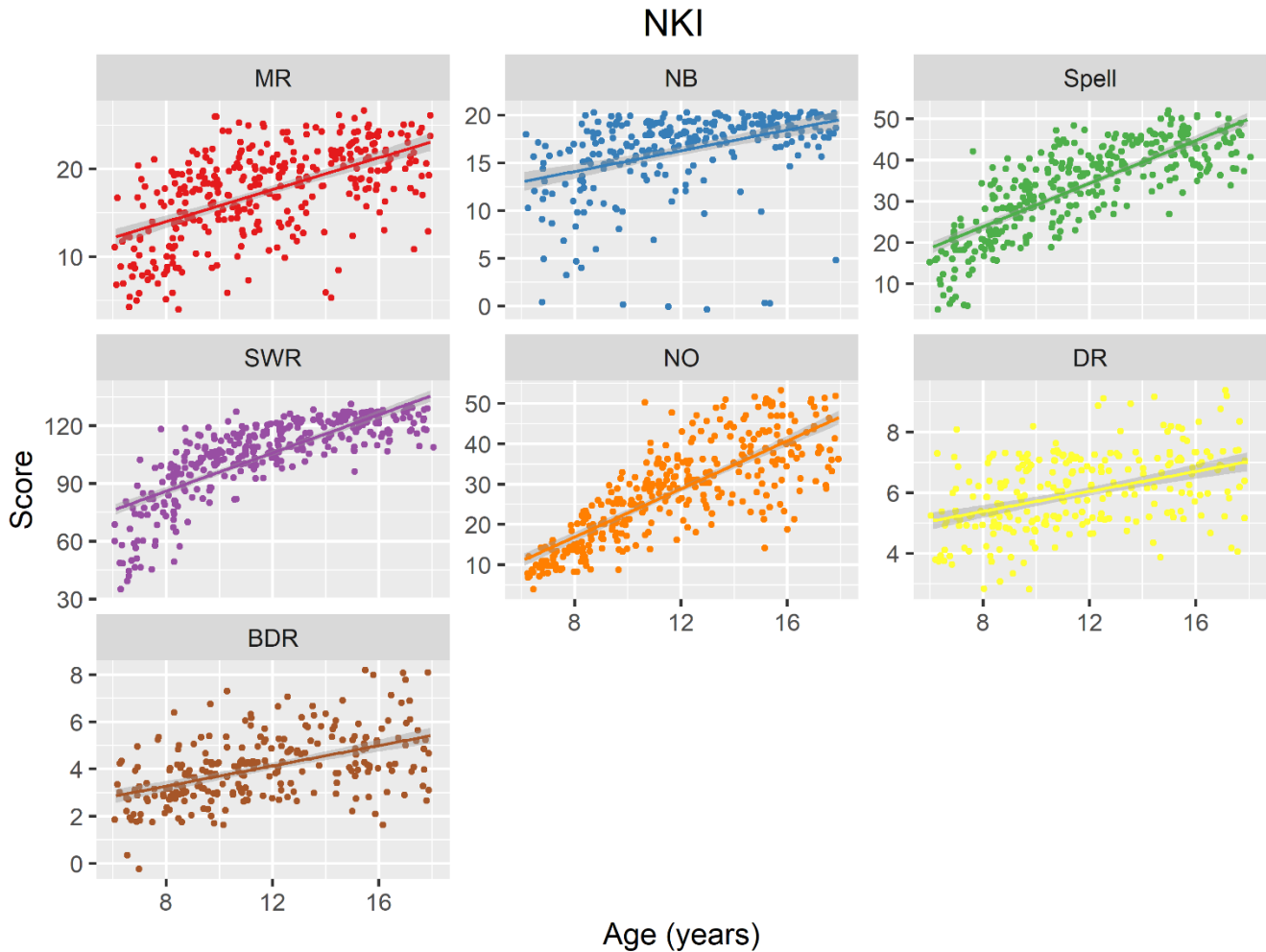


Fig. 2. Scatterplots of cognitive task scores across age for CALM and NKI Rockland samples. Lines reflect linear fit.

206 2.1.4 MRI acquisition

207

208 The CALM sample neuroimaging data were obtained at the MRC Cognition and
209 Brain Sciences Unit, Cambridge, UK. Scans were acquired on the Siemens 3 T Tim Trio
210 system (Siemens Healthcare, Erlangen, Germany) via 32-channel quadrature head coil.
211 Echo-planar diffusion-weighted images were used to attain diffusion scans using a set of 60
212 non-collinear directions and a weighting factor of $b=1000s*mm^{-2}$ combined with a T2-
213 weighted ($b=0$) volume. Whole brain coverage was obtained with 60 contiguous axial slices
214 and an isometric image resolution of 2mm. Total echo time and repetition time were 90ms
215 and 8400ms, respectively.

216 For the NKI sample, participants were also scanned using a Siemens 3 T Tim Trio
217 system. All T1-weighted images were attained via magnetization-prepared rapid gradient
218 echo (MPRAGE) sequence with 1mm isotropic resolution. An isotopic set of gradients using

219 137 directions with a weighting factor of $b=1000s \cdot mm^{-2}$ and an isotropic resolution of 2mm
220 were used to acquire diffusion scans. For further details regarding scan sequences, see
221 http://fcon_1000.projects.nitrc.org/indi/enhanced/mri_protocol.html.

222

223 *2.1.5 White matter connectome construction*

224

225 Note that part of the following pipeline is identical to that described in Bathelt et al.,
226 (in press). Diffusion-weighted images were pre-processed to create a brain mask based on
227 the b0-weighted image (FSL BET; Smith, 2002) and to correct for movement and eddy
228 current-induced distortions (eddy; Graham et al., 2016). Subsequently, the diffusion tensor
229 model was fitted and fractional anisotropy (FA) maps were calculated (dtifit). Images with a
230 between-image displacement greater than 3mm as indicated by FSL eddy were excluded
231 from further analysis. All steps were carried out with FSL v5.0.9 and were implemented in a
232 pipeline using NiPyPe v0.13.0 (Gorgolewski et al., 2011). To extract FA values for major
233 white matter tracts, FA images were registered to the FMRIB58 FA template in MNI space
234 using a sequence of rigid, affine, and symmetric diffeomorphic image registration (SyN) as
235 implemented in ANTS v1.9 (Avants et al., 2008). Visual inspection indicated good image
236 registration for all participants. Subsequently, binary masks from a probabilistic white matter
237 atlas (threshold at >50% probability) in the same space were applied to extract FA values for
238 white matter tracts (see below).

239 Participant movement, particularly in developmental samples, can significantly affect
240 the quality, and, hence, statistical analyses of MRI data. Therefore, we undertook several
241 procedures to ensure adequate MRI data quality and minimize potential biases due to
242 subject movement. First, for the CALM sample, children were trained to lie still inside a
243 realistic mock scanner prior to their actual scans. Secondly, for both samples, all T1-
244 weighted images and FA maps were visually examined by a qualified researcher to remove
245 low quality scans. Lastly, quality of the diffusion-weighted data were evaluated in both
246 samples by calculating the framewise displacement between subsequent volumes in the
247 sequence. Only data with a maximum between-volume displacement below 3mm were
248 included in the analyses. All steps were carried out with FMRIB Software Library v5.0.9 and
249 implemented in the pipeline using NiPyPe v0.13.0 (see
250 <https://nipy.readthedocs.io/en/latest/>).

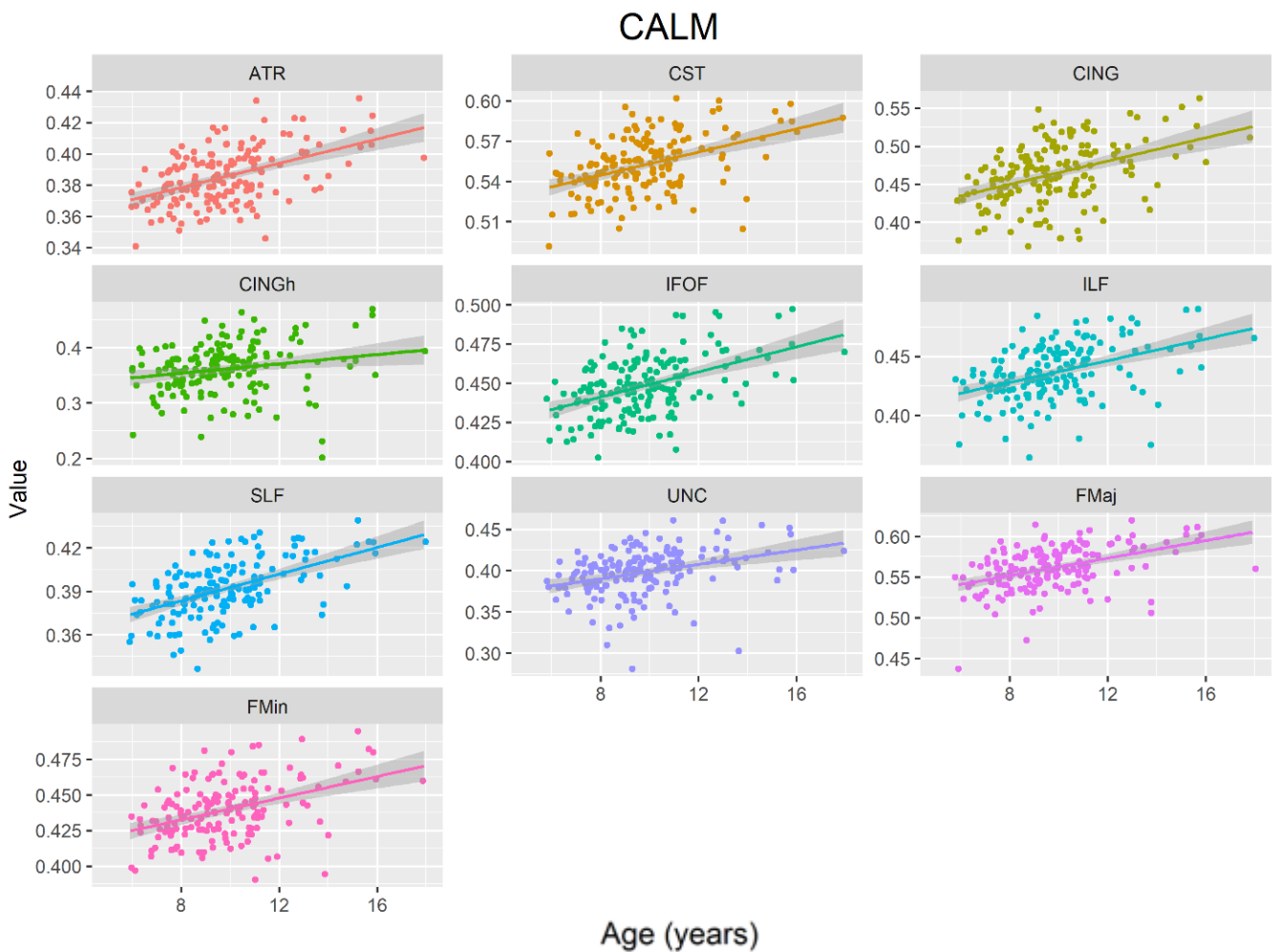
251

252 *2.1.6 Neural measures: white matter and fractional anisotropy*

253

254 To approximate white matter contributions to fluid and crystallized ability, we
255 analyzed fractional anisotropy (FA; see Wandell, 2016). We based our choice of FA on

256 previous studies of white matter in developmental samples (de Mooij et al., 2018; Kievit et
257 al., 2016). We used FA as a general summary metric of white matter microstructure as it
258 cannot directly discern between specific cellular components (e.g. axonal diameter, myelin
259 density, water fraction). Mean FA was computed for 10 bilateral tracts as defined by the
260 Johns Hopkins University DTI-based white matter tractography atlas (see Fig. 1 of Hua et
261 al., 2008): forceps minor (FMin), forceps major (FMaj), anterior thalamic radiations (ATR),
262 cingulate gyrus (CING), superior longitudinal fasciculus (SLF), inferior longitudinal fasciculus
263 (ILF), corticospinal tract (CST), uncinete fasciculus (UNC), cingulum [hippocampus]
264 (CINGh), and inferior fronto-occipital fasciculus (IFOF). Fig. 3 shows the cross-sectional
265 trends of FA across the age range.
266



267

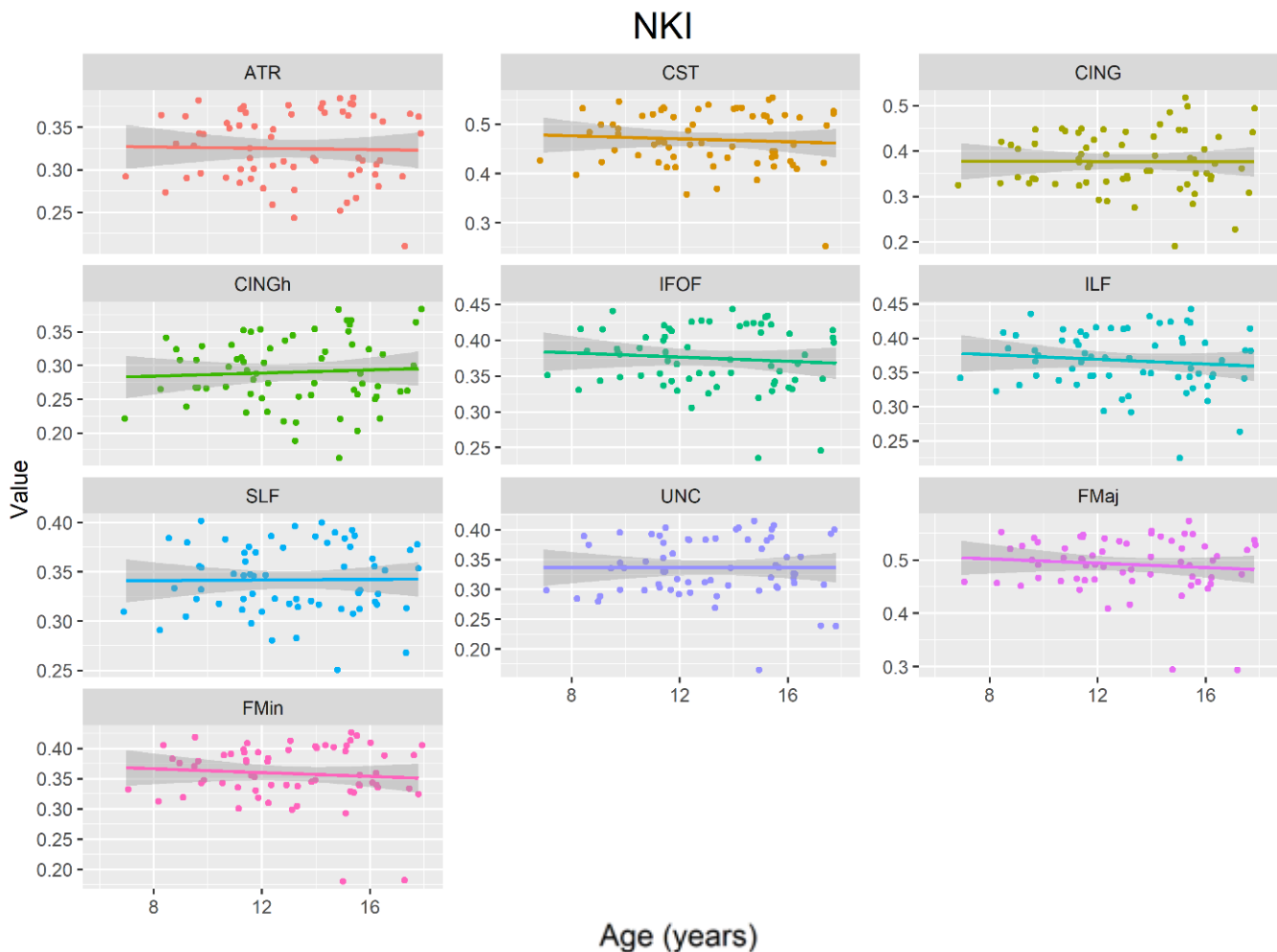


Fig. 3. Scatterplots of FA values for all white matter tracts across age for CALM and NKI Rockland samples. Lines reflect linear fit. Note that the age trends are more pronounced in CALM than in the NKI sample, possibly due to lower sample size in NKI (N=65).

268

269 **3. Results**

270

271 *3.1 Covariance among cognitive abilities cannot be captured by a single-factor*

272

273 In accordance with our [preregistered analysis plan](#), we first describe model fit for the
 274 measurement models of the cognitive data only. First, we tested hypothesis 1: that gc and gf
 275 are separable constructs in childhood and adolescence. More specifically, we tested the
 276 hypothesis that the covariance among scores on cognitive tests would be better captured by
 277 a two-factor (gc-gf) model than a single-factor (e.g. g) model. In support of this prediction,
 278 the single-factor model fit the data poorly: $\chi^2(27) = 317.695$, $p < .001$, RMSEA = .146 [.132
 279 .161], CFI = .908, SRMR = .040, Yuan-Bentler scaling factor = 1.090, suggesting that cognitive
 280 performance was not well represented by a single-factor. The two-factor (gc-gf) model also
 281 displayed poor model fit ($\chi^2(24) = 196.348$, $p < .001$, RMSEA = .119 [.104 .135], CFI = .946,

282 SRMR=.046, Yuan-Bentler scaling factor= 1.087), although it fit significantly better
283 ($\chi^2\Delta=119.41$, $df\Delta=3$, $AIC\Delta=127$, $p<0.001$) than the single-factor model.

284 To investigate the source of poor fit, we examined modification indices (Schermelleh-
285 Engel et al., 2003), which quantify the expected improvement in model fit if a parameter is
286 freed. Modification indices suggested that the Peabody Picture Vocabulary Test had a very
287 strong cross-loading onto the fluid intelligence latent factor. The Peabody Picture Vocabulary
288 Test (PPVT), often considered a crystallized measure in adult populations, asks participants
289 to choose the picture (out of four multiple-choice options) corresponding to the meaning of
290 the word spoken by an examiner. Including a cross-loading between gf and the PPVT
291 drastically improved goodness of fit ($\chi^2\Delta= 67.52$, $df\Delta=1$, $AIC\Delta=100$, $p<0.001$) to adequate
292 ($\chi^2(23)=104.533$, $p<.001$, $RMSEA=.083$ [.067 .099], $CFI=.975$, $SRMR=.025$, Yuan-Bentler
293 scaling factor= 1.069). A likely explanation of this result is that such tasks may draw
294 considerably more on executive, gf-like abilities in younger, lower ability samples. For a
295 more thorough investigation of the loading of PPVT across development, see Supplementary
296 Material. Notably, fitting the PPVT as a *solely* fluid task (i.e. removing it as a measurement of
297 gc entirely) did not significantly decrease model fit ($\chi^2\Delta= 2.058$, $df\Delta=1$, $AIC\Delta=1$, $p=0.152$).
298 Therefore, we decided to proceed with the more parsimonious PPVT gf-only model ($\chi^2(24)$
299 $=106.382$, $p<.001$, $RMSEA=.082$ [.066 .098], $CFI=.972$, $SRMR=.025$, Yuan-Bentler scaling
300 factor= 1.073). We note that although this is a data-driven modification, we believe it would
301 likely generalize to samples with similarly low ages and abilities.

302 Next, we examined whether the single or two-factor model fit best in the NKI sample.
303 The single-factor model fit the data adequately ($\chi^2(14)=41.329$, $p<.001$, $RMSEA=.075$ [.049
304 .102], $CFI=.983$, $SRMR=.029$, Yuan-Bentler scaling factor=.965). Moreover, all loadings
305 between the cognitive tasks and g were significant ($p<.05$) and high (standardized
306 loadings $\geq.5$). However, as was the case in the CALM sample, the two-factor model showed
307 considerably better fit ($\chi^2(12)=19.732$, $p=.072$, $RMSEA=.043$ [.000 .075], $CFI=.995$,
308 $SRMR=.018$, Yuan-Bentler scaling factor=.956) compared to the single-factor model
309 ($\chi^2\Delta=20.661$, $df\Delta=2$, $AIC\Delta=17$, $p<0.001$). It should be noted that, given the differences in
310 tasks measured between the samples, gf and working memory were assumed to be
311 measurements of the same latent factor, rather than separable factors. A similar competing
312 model where gf and working memory were modeled as separate constructs with working
313 memory loaded onto gf, similarly to the best-fitting model for the CALM sample (see Fig. 4),
314 showed comparable model fit and converging conclusions with further analyses. Overall,
315 these findings suggested, that for both the NKI and CALM samples, a two-factor model with
316 separate gc and gf factors provided a better account of individual differences in intelligence
317 than a single-factor model.

318

319 3.2 Evidence of age differentiation between crystallized and fluid ability

320

321 We investigated the relationship between gc and gf in development to see whether
322 we could observe evidence for age differentiation as predicted by hypothesis 2. Age
323 differentiation (e.g. Hülür et al., 2011) would predict decreasing covariance between gc and
324 gf from childhood to adolescence. We fit a multigroup confirmatory factor analysis to assess
325 fit on our younger (N=279) and older (N=272) participant cohorts. The model had acceptable
326 fit ($\chi^2(48) = 142.214$, $p < .001$, RMSEA = .085 [.069 .102], CFI = .960, SRMR = .037, Yuan-
327 Bentler scaling factor = 1.019). However, a likelihood ratio test, showed that model fit did not
328 decrease significantly when imposing equal covariance between gc and gf in the younger
329 and older participant subgroups ($\chi^2\Delta = 0.323$, $df\Delta = 1$, $AIC\Delta = 2$, $p = 0.57$). This suggested no
330 evidence for age differentiation in the CALM sample. However, the lack of association could
331 be due to limitations of using median splits to investigate age differences when independent
332 (or latent in our case) variables are correlated (Iacobucci et al., 2015). For instance, if the
333 age range of differences in behavioral associations between gc and gf lies elsewhere, the
334 median split may not be sensitive enough to detect it. To test this explicitly, we next fit SEM
335 trees (Brandmaier et al., 2013) to the cognitive data.

336 We estimated SEM trees in the CALM sample by specifying the cognitive model with
337 age as a continuous covariate. We observed a SEM tree split at age 9.12, yielding two
338 groups (younger participants = 290, older participants = 261). This split was accompanied by
339 a decrease in the covariance between gc and gf, providing support for age differentiation
340 using a more exploratory approach to determine the optimal age split (SEM tree: 9.12 versus
341 median split: 8.91).

342 Next, as in the CALM cohort, we fit a multigroup model with younger (N=169) and
343 older (N=168) age groups in the NKI sample, which produced good fit ($\chi^2(24) = 33.736$,
344 $p = .089$, RMSEA = .047 [.000 .081], CFI = .991, SRMR = .035, Yuan-Bentler scaling
345 factor = .916). In contrast to CALM, imposing equality constraints on the covariance between
346 gc and gf across age groups revealed a lower gc-gf correlation for the older (.811) compared
347 to the younger participants cohort (1.008). This revealed significant difference in model fit
348 compared to the freely-estimated model ($\chi^2\Delta = 61.244$, $df\Delta = 1$, $AIC\Delta = 46$, $p < 0.001$). This
349 suggested evidence for age differentiation in the NKI sample using multigroup models.

350 In contrast to the multigroup model outcome, the NKI SEM tree model under identical
351 specifications as in CALM failed to produce an age split. A possible explanation is that to
352 penalize for multiple testing we relied on Bonferroni-corrected alpha thresholds for the SEM
353 tree. If, as seems to be the case here, the true split lies (almost) exactly on the median split,
354 then the SEM tree will have slightly less power than conventional multigroup models. These
355 differences between analyses methods suggested that the age differentiation observed here

356 is likely modest in size. Taken together, we interpret our findings as evidence for a small,
357 age-specific but significant decrease in gc-gf covariance in both cohorts, which is compatible
358 with age differentiation such that, for younger participants, gc and gf factors are almost
359 indistinguishable, whereas for older participants a clearer separation emerges.

360

361 *3.3 Violation of metric invariance suggests differences in relationships among* 362 *cognitive abilities in childhood and adolescence*

363

364 Finally, we more closely examined age-related differences in cognitive architecture
365 (e.g. factor loadings) by examining metric invariance (Putnick and Bornstein, 2016). Testing
366 this in the CALM sample as a two-group model by imposing equality constraints on the factor
367 loadings (fully constrained) showed that the freely-estimated model (no factor loading
368 constraints) outperformed the fully-constrained model ($\chi^2\Delta=107.05$, $df\Delta=7$, $AIC\Delta=82$,
369 $p<0.001$), indicating that metric invariance was violated. This violation of metric invariance
370 suggested that the relationship between the cognitive tests and latent variables was different
371 in the two age groups. Closer inspection suggested that the differences in loadings were not
372 uniform, but rather showed a more complex pattern of age-related differences (see Table 2
373 for more details). Some of the most pronounced differences include an increase of the
374 loading of matrix reasoning onto gf as well as increased loading of digit recall and dot matrix
375 onto working memory across age groups.

376 Similarly, in the NKI cohort, the freely-estimated model outperformed the constrained
377 model ($\chi^2\Delta=41.111$, $df\Delta=5$, $AIC\Delta=33$, $p<0.001$), indicating that metric invariance was again
378 violated as in CALM. This suggests that the relationship between the cognitive tests and the
379 latent factors differed across age groups. The pattern of factor loadings differed in some
380 respects from CALM. For example, the loading of the N-back task onto gf showed the largest
381 difference across age groups in the NKI sample. However, as CALM did not include the N-
382 back task, we cannot directly interpret this as a difference between the cohorts. For detailed
383 comparisons among factor loadings between age groups in both samples, refer to Table 2.
384 The overall pattern in both samples suggested small and varied differences in the
385 relationship between the latent factors and observed scores. A plausible explanation is that
386 the same task draws on a different balance of skills as children differ in age and ability. Our
387 findings concerning the latent factors should be interpreted in this light as it seems likely that
388 in addition to age differentiation (and possibly dedifferentiation) effects, the nature of the
389 factors also differed slightly across the age range studied here.

390

Sample	Relationship	Younger Participants	Older Participants	
CALM	gc↔gf	.89	.93	
	gf→WM	.96	.90	
	gf→MR	.59	.74	
	gf→PPVT	.75	.76	
	WM→DR	.56	.68	
	WM→BDR	.76	.79	
	WM→Dot	.58	.67	
	WM→MRX	.59	.56	
	gc→gcV	.89	.79	
	gc→NO	.87	.87	
	gcV→SWR	.94	.91	
	gcV→Spell	.87	.91	
	NKI	gc↔gfWM	1	.81
		gc→gcV	.96	.87
gfWM→MR		.69	.60	
gfWM→DR		.38	.54	
gfWM→BDR		.50	.53	
gfWM→NB		.55	.35	
gc→NO		.90	.76	
gcV→SWR		.93	.89	
gcV→Spell		.97	.88	

Table 2. Standardized Path Estimates for Cognitive Assessments in CALM & NKI Rockland samples. Note that age groups were determined according to the median split (CALM: 8.91 years, NKI: 11.38 years)

391

392 *3.4 The neural architecture of gc and gf indicates unique contributions of multiple*
 393 *white matter tracts to cognitive ability*

394

395 We next focused on the white matter regression coefficients to inspect the neural
 396 underpinnings of gc and gf. In line with hypothesis 3, we wanted to explore whether
 397 individual white matter tracts made independent contributions to gc and gf. First, we
 398 examined whether a single-factor model could account for covariance in white matter
 399 microstructure across our ten tracts. If so, then scores on such a latent factor would
 400 represent a parsimonious summary for neural integrity. However, this model showed poor fit
 401 ($\chi^2(35) = 124.810$, $p < .001$, $RMSEA = .132$ [.107 .157], $CFI = .938$, $SRMR = .039$, Yuan-Bentler
 402 scaling factor = 1.114), suggesting separate influences from white matter regions in
 403 supporting cognitive abilities. To examine whether the white matter tracts showed specific
 404 and complementary associations with cognitive performance, we fit a MIMIC model. Doing
 405 so, we observed that 5 out of the 10 tracts showed significant relations with gc and/or gf (Fig.

406 4). Specifically, the anterior thalamic radiations, forceps major, and forceps minor had
407 moderate to strong associations with gc with similar relations seen for gf for the superior
408 longitudinal fasciculus, forceps major, and the cingulate gyrus. Interestingly, the forceps
409 minor exhibited a negative association with gf. This could be due to modeling several highly
410 correlated paths simultaneously since this relationship was not found when only the forceps
411 minor was modeled onto gc (standardized estimate=.426) and gf (standardized
412 estimate=.386, see Tu et al., 2008). Together, individual differences in white matter
413 microstructure explained 32.9% in crystallized and 33.6 % in fluid ability.

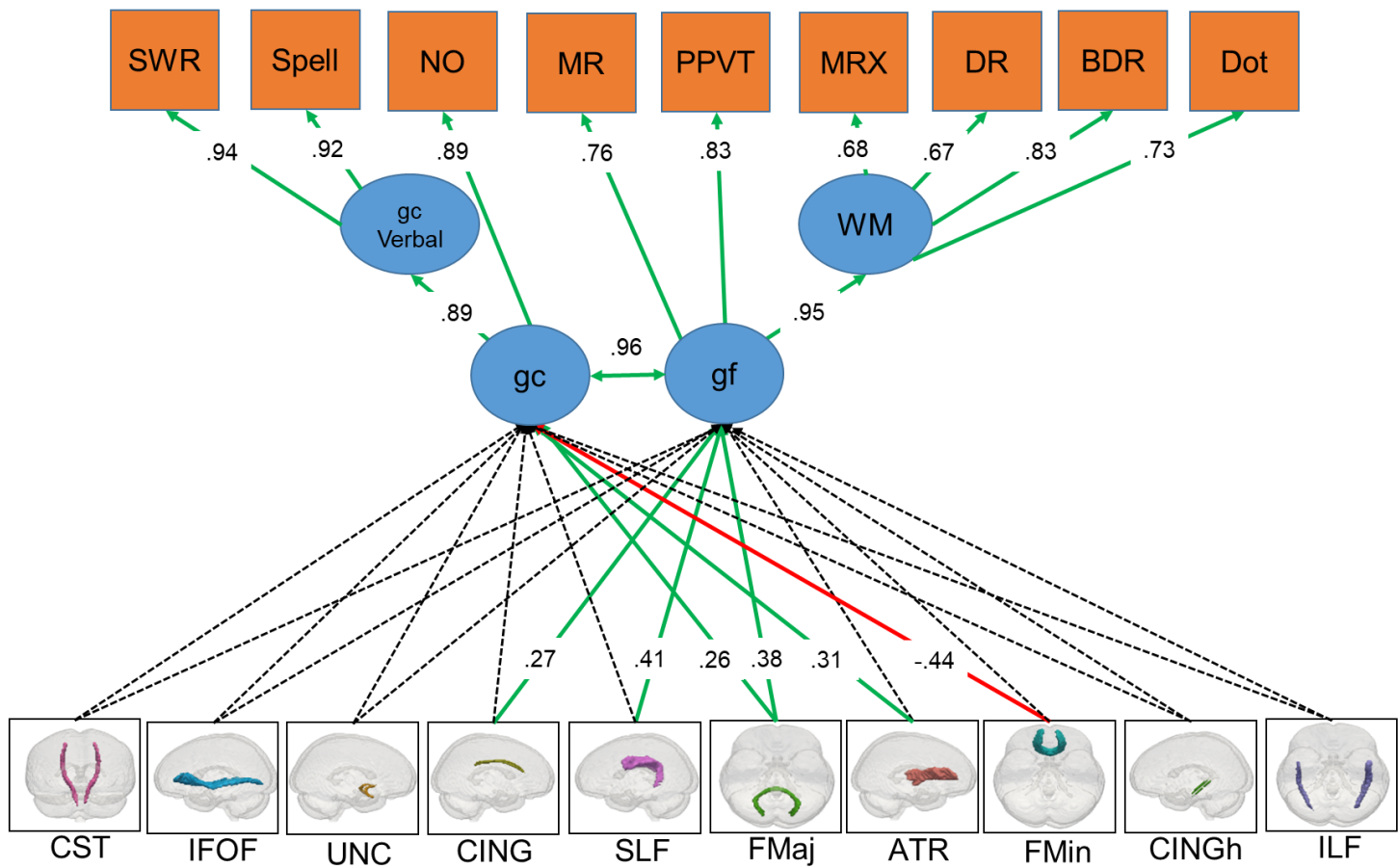
414 As in the CALM sample, the single-factor white matter model produced poor fit (χ^2
415 (35) =131.637, $p < .001$, RMSEA=.201 [.165 .238], CFI= .924, SRMR=.023, Yuan-Bentler
416 scaling factor=.950) in the NKI sample. Therefore, we fit a multi-tract MIMIC model. The
417 superior longitudinal fasciculus emerged as the only tract to significantly load onto gc or gf
418 (Fig. 4). This result was likely due to lower power associated with a small subset of
419 individuals with white matter data (N=65, see Discussion for further investigation). In NKI,
420 the same set of tracts explained 29.7% and 26.7% of the variance in gc and gf, respectively,
421 yielding similar joint effect sizes as in the CALM sample. Together, these findings
422 demonstrated generally similar associations between white matter microstructure and
423 cognitive abilities in the CALM and NKI samples. Therefore, it seems to be the case that, in
424 both typically and atypically (struggling learners) developing children and adolescents,
425 individual white matter tracts make distinct contributions to crystallized and fluid ability.

426

CALM

Model fit: $\chi^2(94) = 196.98, p < .001$; RMSEA = .045
[.036 .053]; CFI = .971; SRMR = .028

-----> non-significant
-----> p<.05 positive
-----> p<.05 negative



427

NKI

Model fit: $\chi^2(62) = 83.42$, $p < .05$; RMSEA = .032 [.009
.048]; CFI = .987; SRMR = .107

—→ non-significant
—→ p<.05 positive
—→ p<.05 negative

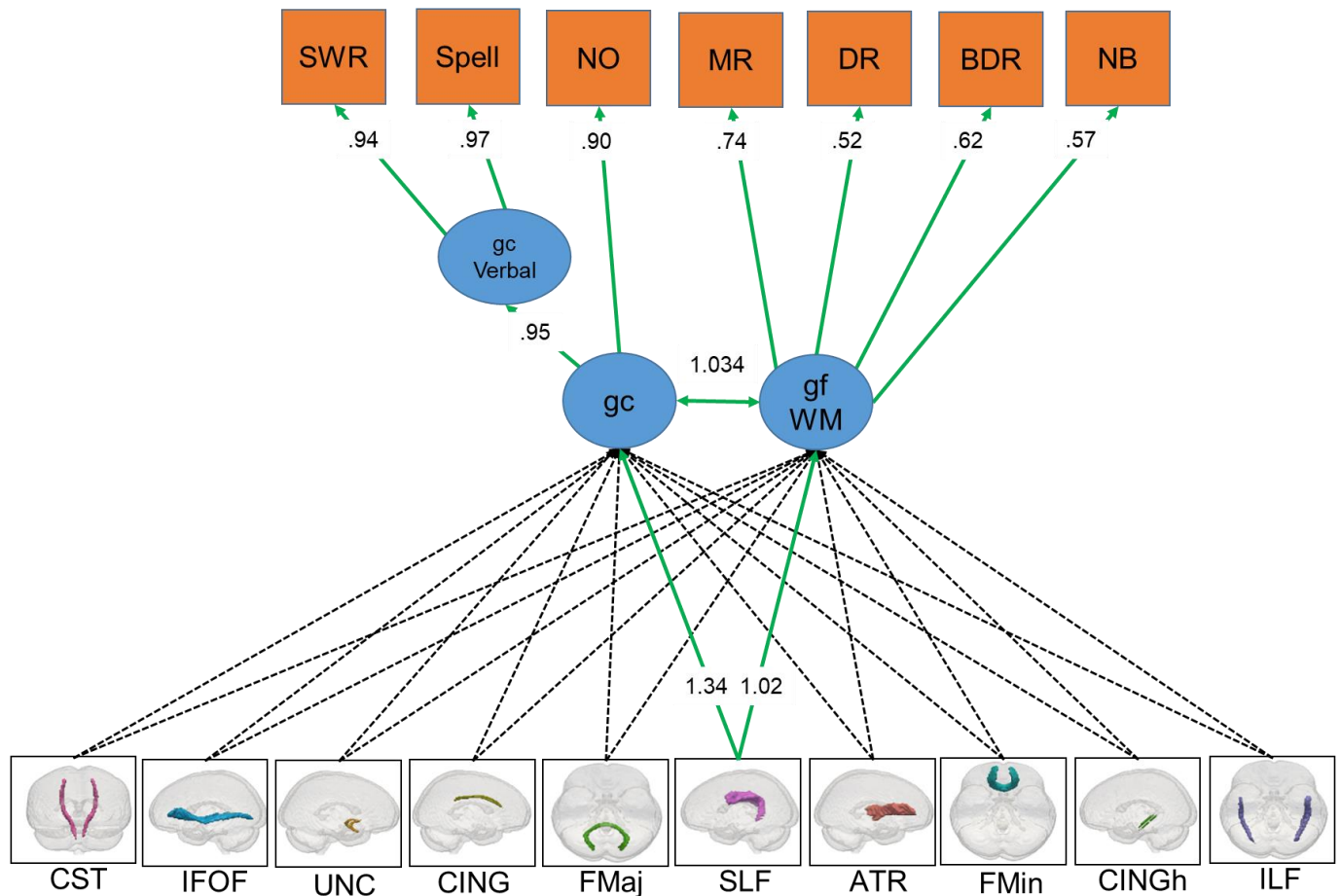


Fig. 4. MIMIC models displaying standardized parameter estimates and regression coefficients for all cognitive measures and white matter tracts for complete CALM and NKI Rockland samples. Note that the greater than 1 standardized factor loadings in NKI may occur in the presence of highly-correlated factors (Jöreskog, 1999)

428 3.5 Support for a neurocognitive account of the age differentiation-dedifferentiation 429 hypothesis

430

431 Lastly, to address our fourth and final preregistered hypothesis, we examined
432 whether brain-behavior associations differed across the developmental age range. We
433 hypothesized that the relationship between the white matter tracts and cognitive abilities
434 would decrease across the age range, in support of the differentiation hypothesis. Using a
435 multigroup model, we compared the strength of brain-behavior relationships between
436 younger and older participants to test whether white matter contributions to gc and gf

437 differed in development. Contrary to our prediction, we observed that, in the CALM sample,
 438 a freely estimated model, where the brain-behavior relationships were allowed to vary across
 439 age groups, did not outperform the constrained model ($\chi^2 \Delta = 12.16$, $df\Delta = 10$, $AIC\Delta = 9$,
 440 $p = 0.27$). This suggested that the contributions of white matter tracts did not vary significantly
 441 between age groups when examined using multigroup models.

442 As before, we estimated a SEM tree model. In contrast to the multigroup model, we
 443 observed that multiple white matter tracts *did* differ in their associations with gc and/or gf.
 444 These differences manifested in different ways for gc and gf. For example, the correlations
 445 between the cingulum, superior longitudinal fasciculus, and forceps major and gf decreased
 446 with increasing age, in line with age differentiation. On the other hand, the forceps major,
 447 forceps minor and anterior thalamic radiations demonstrated a more complicated pattern
 448 with each tract displaying two age splits. For the first split (around age 8), the regression
 449 strength decreased before spiking again around age 11 (Table 3, also see Fuhrmann et al.,
 450 2019). Given that all first splits showed a decrease between white matter and cognition, and
 451 all second splits revealed an increase compared to the first, this suggests a non-monotonic
 452 pattern of brain-behavior reorganization that cannot be fully captured by age differentiation
 453 or dedifferentiation (Hartung et al., 2018) but may be in line with theories such as Interactive
 454 Specialization (Johnson, 2011), which provides a range of mechanisms which may induce
 455 age-varying brain-behavior strengths.

456 Lastly, we performed the same multigroup analysis for the NKI MIMIC model, but it
 457 failed to converge or produce an age split, likely due to sparsity of the neural data ($N = 65$).
 458 Therefore, this analysis could not be used to replicate the cutoff age used for multigroup
 459 analyses (11.38 years) based on the median split. Further inspection of the only significantly
 460 associated tract, the superior longitudinal fasciculus, revealed the same trend for gc and gf
 461 with decreased correlations with increasing age (Table 3). Overall, our findings suggest the
 462 need for a neurocognitive account of age differentiation-dedifferentiation from childhood into
 463 adolescence.

464

	<i>Relationship</i>	<i>Estimate Before Split</i>	<i>Age of Split 1</i>	<i>Estimate After Split</i>	<i>Age of Split 2</i>	<i>Estimate After Split</i>
CALM	gc↔gf	.64	9.12	.59	NS	NS
	gf→CING	.29	7.38	.18	NS	NS
	gf→SLF	.38	7.38	.29	NS	NS
	gf→FMaj	.38	7.38	.26	NS	NS
	gc→FMaj	.24	8.29	.04	10.79	.42
	gc→ATR	.30	7.62	.13	10.79	.37
	gc→FMin	-.34	7.62	-.52	10.79	-.25

NKI	gc↔gfWMM	.96	NS	NS	NS	NS
	gf→SLF	.35	13.16	.21	NS	NS
	gc→SLF	.91	9.85	.69	NS	NS

Table 3. SEM tree Results for CALM & NKI Rockland samples. Note: values listed represent unstandardized estimates. NS= no split

465 **4. Discussion**

466

467 *4.1 Summary of findings*

468

469 In this [preregistered](#) analysis, we examined the cognitive architecture as well as the
 470 white matter substrates of fluid and crystallized intelligence in children and adolescents in
 471 two developmental samples (CALM and NKI). Analyses in both samples indicated that
 472 individual differences in intelligence were better captured by two separate but highly
 473 correlated factors (gc and gf) of cognitive ability as opposed to a single global factor (*g*).
 474 Further analysis suggested that the covariance between these factors decreased slightly
 475 from childhood to adolescence, in line with the age differentiation hypothesis of cognitive
 476 abilities (Garrett, 1946; Hülür et al., 2011).

477 We observed multiple, partially independent contributions of specific tracts to
 478 individual differences in gc and gf. The clearest associations were observed for the anterior
 479 thalamic radiations, cingulum, forceps major, forceps minor, and superior longitudinal
 480 fasciculus, all of which have been implicated to play a role in cognitive functioning in
 481 childhood and adolescence (Krogsrud et al., 2018; Navas-Sánchez et al., 2014; Peters et al.,
 482 2014; Tamnes et al., 2010; Urger et al., 2015; Vollmer et al., 2017). However, except for the
 483 superior longitudinal fasciculus, these tracts were not significant in NKI Rockland sample. A
 484 possible explanation for this is the difference in imaging sample size between the cohorts
 485 (N=165 in the CALM sample versus N=65 in the NKI Rockland sample). This difference
 486 implies sizeable differences in power (73.4% in CALM versus 36.2% in NKI, assuming a
 487 standardized effect size of 0.2) to identify weaker individual pathways.

488 The most consistent association, observed in both samples, was between the
 489 superior longitudinal fasciculus, a region known to be important for language and cognition,
 490 which significantly contributed to cognitive ability in both CALM (gf only) and NKI (gc and gf).
 491 The superior longitudinal fasciculus is a long myelinated bidirectional association fiber
 492 pathway that runs from anterior to posterior cortical regions and through the major lobes of
 493 each hemisphere (Kamali et al., 2014), and has been associated with memory, attention,
 494 language, and executive function in childhood and adolescence in both healthy and atypical
 495 populations (Frye et al., 2010; Urger et al., 2015). Therefore, given its widespread links

496 throughout the brain, which include temporal and fronto-parietal regions, it is no surprise that
497 it was found to be significantly related to both gc and gf in our samples.

498 Together, these results are in line with previous research relating fractional
499 anisotropy (FA) and cognitive ability. For instance, Peters et al., 2014 found that age-related
500 differences in cingulum FA mediated differences in executive functioning. Moreover, white
501 matter changes in the forceps major have been linked to higher performance on working
502 memory tasks (Krogsrud et al., 2018). The remaining tracts (superior longitudinal fasciculus
503 and anterior thalamic radiations) have also been positively correlated with verbal and non-
504 verbal cognitive performance in childhood and adolescence (Tamnes et al., 2010; Urger et
505 al., 2015). We also observed more surprising negative pathways, such as between gc and
506 the forceps minor in the CALM sample. However, closer inspection showed that the simple
507 association between forceps minor and gc was *positive*, suggesting the negative pathway is
508 likely the consequence of the simultaneous inclusion of collinear predictors (see Tu et al.,
509 2008).

510 Finally, using SEM trees (Brandmaier et al., 2013), we observed that white matter
511 contributions to gc and gf differed between participants of different ages. In CALM, the
512 contributions of the cingulum, superior longitudinal fasciculus, and forceps major weakened
513 with increasing age for gf. For gc, however, the forceps major and forceps minor, and the
514 anterior thalamic radiations exhibited a more complex pattern with each tract providing
515 significantly different effects on crystallized intelligence at two distinct time points in
516 development. In NKI, the superior longitudinal fasciculus became less associated with both
517 gc and gf. Considering that decreases in white matter relations to gc and gf occurred before
518 covariance decreases found between gc and gf suggest that differences in white matter
519 development may underlie subsequent individual differences in cognition. In a related project
520 (Fuhrmann et al., 2019, Table 6) we observed age-related differences in associations
521 despite focusing on different cognitive factors (processing speed and working memory).

522 Overall, our findings align with a neurocognitive interpretation of age differentiation-
523 dedifferentiation hypothesis, which would predict that cognitive abilities and their neural
524 substrates become more differentiated (less correlated) until the onset of maturity, followed
525 by an increase (dedifferentiation) in relation to each other until late adulthood (Hartung et al.,
526 2018). However, we note that the evidence for age differentiation-dedifferentiation was not
527 always robust across analyses methods or samples, suggesting only small effect sizes.

528

529 *4.2 Limitations of the present study*

530 First and foremost, all findings here were observed in cross-sectional samples. To
531 better understand effects such as age differentiation and dedifferentiation, future studies will
532 need to model age-related changes within the same individual. The complexity and expense

533 of collecting such longitudinal data has long precluded such investigations, but new cohorts
534 such as the ABCD sample (Volkow et al., 2018) will allow us to model longitudinal changes
535 in the future. Secondly, although the majority of our findings are similar across our cohorts,
536 some differences were observed, particularly in white matter effects. This may reflect
537 statistical variability, differences in sample size and associated differences in power, or true
538 differences between samples. Moreover, the white matter differences observed could also
539 be due to the scans being obtained at different scanner sites, although this is unlikely to
540 have produced considerable differences for all raw images were processed using the same
541 pipeline, and previous work suggests that FA is quite a robust measure in multi-site
542 comparison (see Vollmar et al., 2010).

543 CALM consists of children with referrals for any difficulties related to learning,
544 attention or memory (Holmes et al., 2019). The NKI Rockland sample, in contrast, is a
545 United States population representative sample (Nooner et al., 2012). Both samples are
546 composed of large cohorts that underwent extensive phenotyping and population-specific
547 representative sampling. Therefore, we argue that our results generalize to 'typical' and
548 'atypical' samples of neurocognitive development.

549

550 *4.3 Conclusions*

551

552 The present analyses revealed that crystallized and fluid intelligence factors
553 explained a significant amount of variance in test performance in two large child and
554 adolescent samples. These results were found in both typically and atypically (struggling
555 learners) developing cohorts, demonstrating the generalized notion that cognitive ability is
556 better understood as a two-factor rather than a single-factor phenomenon in childhood and
557 adolescence. The addition of white matter microstructure indicated independent
558 contributions from specific white matter tracts known to be involved in cognitive ability.
559 Moreover, further analyses suggested that the associations between neural and behavioral
560 measures differed during development.

561 Overall, these results support a neurocognitive age differentiation-dedifferentiation
562 hypothesis of cognitive abilities whereby the relation between white matter and cognition
563 become more differentiated (less correlated) in pre-puberty and then dedifferentiate (become
564 more correlated) during early puberty. However, despite our use of novel and more sensitive
565 statistical methods (SEM trees), the samples used were cross-sectional and, therefore, are
566 not adequate to make causal claims about the neurocognitive dynamics of intelligence in
567 childhood and adolescence. Future studies should take this limitation into account when
568 designing experiments attempting to clarify such statements.

569

570 **Declarations of interest**

571

572 None.

573

574 **Acknowledgements**

575

576 The Centre for Attention Learning and 639 Memory (CALM) research clinic is based
577 at and supported by funding from the MRC Cognition and Brain Sciences Unit, University of
578 Cambridge. The Principal Investigators are Joni Holmes (Head of CALM), Susan Gathercole
579 (Chair of CALM Management Committee), Duncan Astle, Tom Manly and Rogier Kievit. Data
580 collection is assisted by a team of researchers and PhD students at the CBU that includes
581 Sarah Bishop, Annie Bryant, Sally Butterfield, Fanchea Daily, Laura Forde, Erin Hawkins,
582 Sinead O'Brien, Cliodhna O'Leary, Joseph Rennie, and Mengya Zhang. The authors wish to
583 thank the many professionals working in children's services in the South-East and East of
584 England for their support, and to the children and their families for giving up their time to visit
585 the clinic. We would also like to thank all NKI-RS participants and researchers. I.L.S.-K. is
586 supported by the Cambridge Trust. D.F., J.B. and G.S. Borgeest are supported by the UK
587 Medical Research Council (MRC). J.A. is funded by the Studienstiftung des deutschen
588 Volkes (German Academic Scholarship Foundation). This project also received funding from
589 the European Union's Horizon 2020 research and innovation programme (grant agreement
590 number 732592). R. A. Kievit is supported by the Wellcome Trust (Grant No. 107392/Z/15/Z)
591 and the UK Medical Research Council SUAG/014 RG91365.

592

593 **References**

594

- 595 Alloway, T.P., 2007. Automated Working Memory Assessment (AWMA).
596 Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image
597 registration with cross-correlation: Evaluating automated labeling of elderly and
598 neurodegenerative brain. *Medical Image Analysis, Special Issue on The Third
599 International Workshop on Biomedical Image Registration – WBIR 2006* 12, 26–41.
600 <https://doi.org/10.1016/j.media.2007.06.004>
601 Bathelt, J., Zhang, M., Johnson, A., Astle, D., (in press). The cingulum as a marker of
602 individual differences in neurocognitive development. *Scientific Reports*.
603 <https://doi.org/10.17863/CAM.36167>
604 Bickley, P.G., Keith, T.Z., Wolfle, L.M., 1995. The three-stratum theory of cognitive abilities:
605 Test of the structure of intelligence across the life span. *Intelligence* 20, 309–328.
606 [https://doi.org/10.1016/0160-2896\(95\)90013-6](https://doi.org/10.1016/0160-2896(95)90013-6)
607 Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R.,
608 Kenny, S., Bates, T., Mehta, P., Fox, J., 2011. OpenMx: An Open Source Extended
609 Structural Equation Modeling Framework. *Psychometrika* 76, 306–317.
610 <https://doi.org/10.1007/s11336-010-9200-6>
611 Bozdogan, H., 1987. Model selection and Akaike's Information Criterion (AIC): The general
612 theory and its analytical extensions. *Psychometrika* 52, 345–370.
613 <https://doi.org/10.1007/BF02294361>

- 614 Brandmaier, A.M., von Oertzen, T., McArdle, J.J., Lindenberger, U., 2013. Structural
615 equation model trees. *Psychological Methods* 18, 71–86.
616 <https://doi.org/10.1037/a0030001>
- 617 Calvin, C.M., Deary, I.J., Fenton, C., Roberts, B.A., Der, G., Leckenby, N., Batty, G.D., 2011.
618 Intelligence in youth and all-cause-mortality: systematic review with meta-analysis.
619 *International Journal of Epidemiology* 40, 626–644. <https://doi.org/10.1093/ije/dyq190>
- 620 Cattell, R.B., 1967. The theory of fluid and crystallized general intelligence checked at the 5-
621 6 year-old level. *British Journal of Educational Psychology* 37, 209–224.
622 <https://doi.org/10.1111/j.2044-8279.1967.tb01930.x>
- 623 de Mooij, S.M.M., Henson, R.N.A., Waldorp, L.J., Kievit, R.A., 2018. Age Differentiation
624 within Gray Matter, White Matter, and between Memory and White Matter in an Adult
625 Life Span Cohort. *The Journal of Neuroscience* 38, 5826–5836.
626 <https://doi.org/10.1523/JNEUROSCI.1627-17.2018>
- 627 Deary, I.J., Penke, L., Johnson, W., 2010. The neuroscience of human intelligence
628 differences. *Nature Reviews Neuroscience* 11, 201–211.
629 <https://doi.org/10.1038/nrn2793>
- 630 Deary, I.J., Strand, S., Smith, P., Fernandes, C., 2007. Intelligence and educational
631 achievement. *Intelligence* 35, 13–21. <https://doi.org/10.1016/j.intell.2006.02.001>
- 632 Dunn, L.M., Dunn, D.M., 2007. PPVT-4: Peabody picture vocabulary test.
- 633 French, B.F., Finch, W.H., 2008. Multigroup Confirmatory Factor Analysis: Locating the
634 Invariant Referent Sets. *Structural Equation Modeling: A Multidisciplinary Journal* 15,
635 96–113. <https://doi.org/10.1080/10705510701758349>
- 636 Frye, R.E., Hasan, K., Malmberg, B., Desouza, L., Swank, P., Smith, K., Landry, S., 2010.
637 Superior longitudinal fasciculus and cognitive dysfunction in adolescents born
638 preterm and at term: Superior Longitudinal Fasciculus and Cognitive Deficits.
639 *Developmental Medicine & Child Neurology* 52, 760–766.
640 <https://doi.org/10.1111/j.1469-8749.2010.03633.x>
- 641 Fuhrmann, D., Simpson-Kent, I.L., Bathelt, J., Kievit, R.A., 2019. A hierarchical watershed
642 model of fluid intelligence in childhood and adolescence: Supplementary Material.
643 bioRxiv. <https://doi.org/10.1101/435719>
- 644 Garrett, H.E., 1946. A developmental theory of intelligence. *The American Psychologist* 1,
645 372–378. <http://dx.doi.org.ezp.lib.cam.ac.uk/10.1037/h0056380>
- 646 Gignac, G.E., 2014. Dynamic mutualism versus g factor theory: An empirical test.
647 *Intelligence* 42, 89–97. <https://doi.org/10.1016/j.intell.2013.11.004>
- 648 Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L.,
649 Ghosh, S.S., 2011. Nipype: A Flexible, Lightweight and Extensible Neuroimaging
650 Data Processing Framework in Python. *Front. Neuroinform.* 5.
651 <https://doi.org/10.3389/fninf.2011.00013>
- 652 Graham, M.S., Drobnyak, I., Zhang, H., 2016. Realistic simulation of artefacts in diffusion
653 MRI for validating post-processing correction techniques. *NeuroImage* 125, 1079–
654 1094. <https://doi.org/10.1016/j.neuroimage.2015.11.006>
- 655 Gur, R.C., Richard, J., Hughett, P., Calkins, M.E., Macy, L., Bilker, W.B., Brensinger, C.,
656 Gur, R.E., 2010. A cognitive neuroscience-based computerized battery for efficient
657 measurement of individual differences: Standardization and initial construct
658 validation. *Journal of Neuroscience Methods* 187, 254–262.
659 <https://doi.org/10.1016/j.jneumeth.2009.11.017>
- 660 Hartung, J., Doebler, P., Schroeders, U., Wilhelm, O., 2018. Dedifferentiation and
661 differentiation of intelligence in adults across age and years of education. *Intelligence*
662 69, 37–49. <https://doi.org/10.1016/j.intell.2018.04.003>
- 663 Holmes, J., Bryant, A., Gathercole, S.E., the CALM Team, 2019. Protocol for a
664 transdiagnostic study of children with problems of attention, learning and memory
665 (CALM). *BMC Pediatrics* 19. <https://doi.org/10.1186/s12887-018-1385-3>
- 666 Horn, J.L., Cattell, R.B., 1967. Age differences in fluid and crystallized intelligence. *Acta*
667 *Psychologica* 26, 107–129. [https://doi.org/10.1016/0001-6918\(67\)90011-X](https://doi.org/10.1016/0001-6918(67)90011-X)

- 668 Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D.S., Calabresi, P.A., Pekar, J.J.,
669 van Zijl, P.C.M., Mori, S., 2008. Tract probability maps in stereotaxic spaces:
670 Analyses of white matter anatomy and tract-specific quantification. *NeuroImage* 39,
671 336–347. <https://doi.org/10.1016/j.neuroimage.2007.07.053>
- 672 Hülür, G., Wilhelm, O., Robitzsch, A., 2011. Intelligence Differentiation in Early Childhood.
673 *Journal of Individual Differences* 32, 170–179. [https://doi.org/10.1027/1614-](https://doi.org/10.1027/1614-0001/a000049)
674 [0001/a000049](https://doi.org/10.1027/1614-0001/a000049)
- 675 Iacobucci, D., Posavac, S.S., Kardes, F.R., Schneider, M.J., Popovich, D.L., 2015. The
676 median split: Robust, refined, and revived. *Journal of Consumer Psychology* 25,
677 690–704. <https://doi.org/10.1016/j.jcps.2015.06.014>
- 678 Johnson, M.H., 2011. Interactive Specialization: A domain-general framework for human
679 functional brain development? *Developmental Cognitive Neuroscience* 1, 7–21.
680 <https://doi.org/10.1016/j.dcn.2010.07.003>
- 681 Jöreskog, K.G., 1999. How Large Can a Standardized Coefficient be?
682 Jöreskog, K.G., Goldberger, A.S., 1975. Estimation of a Model with Multiple Indicators and
683 Multiple Causes of a Single Latent Variable. *Journal of the American Statistical*
684 *Association* 70, 631–639.
- 685 Juan-Espinosa, M., García, L.F., Colom, R., Abad, F.J., 2000. Testing the age related
686 differentiation hypothesis through the Wechsler's scales. *Personality and Individual*
687 *Differences* 29, 1069–1075. [https://doi.org/10.1016/S0191-8869\(99\)00254-8](https://doi.org/10.1016/S0191-8869(99)00254-8)
- 688 Kamali, A., Sair, H.I., Radmanesh, A., Hasan, K.M., 2014. Decoding the superior parietal
689 lobule connections of the superior longitudinal fasciculus/arcuate fasciculus in the
690 human brain. *Neuroscience* 277, 577–583.
691 <https://doi.org/10.1016/j.neuroscience.2014.07.035>
- 692 Kaufman, A.S., 1975. Factor analysis of the WISC-R at 11 age levels between 6 1/2 and 16
693 1/2 years. *Journal of Consulting and Clinical Psychology* 43, 135–147.
- 694 Kievit, R.A., Davis, S.W., Griffiths, J., Correia, M.M., Cam-CAN, Henson, R.N., 2016. A
695 watershed model of individual differences in fluid intelligence. *Neuropsychologia* 91,
696 186–198. <https://doi.org/10.1016/j.neuropsychologia.2016.08.008>
- 697 Kievit, R.A., van Rooijen, H., Wicherts, J.M., Waldorp, L.J., Kan, K.-J., Scholte, H.S.,
698 Borsboom, D., 2012. Intelligence and the brain: A model-based approach. *Cognitive*
699 *Neuroscience* 3, 89–97. <https://doi.org/10.1080/17588928.2011.628383>
- 700 Krogsrud, S.K., Fjell, A.M., Tamnes, C.K., Grydeland, H., Due-Tønnessen, P., Bjørnerud, A.,
701 Sampaio-Baptista, C., Andersson, J., Johansen-Berg, H., Walhovd, K.B., 2018.
702 Development of white matter microstructure in relation to verbal and visuospatial
703 working memory—A longitudinal study. *PLOS ONE* 13, e0195540.
704 <https://doi.org/10.1371/journal.pone.0195540>
- 705 McArdle, J.J., Hamagami, F., Meredith, W., Bradway, K.P., 2000. Modeling the dynamic
706 hypotheses of Gf–Gc theory using longitudinal life-span data. *Learning and Individual*
707 *Differences* 12, 53–79. [https://doi.org/10.1016/S1041-6080\(00\)00036-4](https://doi.org/10.1016/S1041-6080(00)00036-4)
- 708 Navas-Sánchez, F.J., Alemán-Gómez, Y., Sánchez-Gonzalez, J., Guzmán-De-Villoria, J.A.,
709 Franco, C., Robles, O., Arango, C., Desco, M., 2014. White matter microstructure
710 correlates of mathematical giftedness and intelligence quotient: White Matter
711 Microstructure. *Human Brain Mapping* 35, 2619–2631.
712 <https://doi.org/10.1002/hbm.22355>
- 713 Nooner, K.B., Colcombe, S.J., Tobe, R.H., Mennes, M., Benedict, M.M., Moreno, A.L.,
714 Panek, L.J., Brown, S., Zavitz, S.T., Li, Q., Sikka, S., Gutman, D., Bangaru, S.,
715 Schlachter, R.T., Kamiel, S.M., Anwar, A.R., Hinz, C.M., Kaplan, M.S., Rachlin, A.B.,
716 Adelsberg, S., Cheung, B., Khanuja, R., Yan, C., Craddock, C.C., Calhoun, V.,
717 Courtney, W., King, M., Wood, D., Cox, C.L., Kelly, A.M.C., Di Martino, A., Petkova,
718 E., Reiss, P.T., Duan, N., Thomsen, D., Biswal, B., Coffey, B., Hoptman, M.J., Javitt,
719 D.C., Pomara, N., Sidtis, J.J., Koplewicz, H.S., Castellanos, F.X., Leventhal, B.L.,
720 Milham, M.P., 2012. The NKI-Rockland Sample: A Model for Accelerating the Pace
721 of Discovery Science in Psychiatry. *Frontiers in Neuroscience* 6.
722 <https://doi.org/10.3389/fnins.2012.00152>

- 723 Peters, B.D., Ikuta, T., DeRosse, P., John, M., Burdick, K.E., Gruner, P., Prendergast, D.M.,
724 Szeszko, P.R., Malhotra, A.K., 2014. Age-Related Differences in White Matter Tract
725 Microstructure Are Associated with Cognitive Performance from Childhood to
726 Adulthood. *Biological Psychiatry* 75, 248–256.
727 <https://doi.org/10.1016/j.biopsych.2013.05.020>
- 728 Putnick, D.L., Bornstein, M.H., 2016. Measurement invariance conventions and reporting:
729 The state of the art and future directions for psychological research. *Developmental*
730 *Review* 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- 731 R Core Team, 2018. R: A Language and Environment for Statistical Computing. R
732 Foundation for Statistical Computing, Vienna.
- 733 Rosseel, Y., 2012. lavaan: An R Package for Structural Equation Modeling. *Journal of*
734 *Statistical Software* 48. <https://doi.org/10.18637/jss.v048.i02>
- 735 Schaie, K.W., 1994. The course of adult intellectual development. *American Psychologist* 49,
736 304–313. <https://doi.org/10.1037//0003-066X.49.4.304>
- 737 Schermelleh-Engel, K., Moosbrugger, H., Müller, H., 2003. Evaluating the Fit of Structural
738 Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures.
739 *Methods of Psychological Research* 23–74.
- 740 Schreiber, J.B., Nora, A., Stage, F.K., Barlow, E.A., King, J., 2006. Reporting Structural
741 Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal*
742 *of Educational Research* 99, 323–338. <https://doi.org/10.3200/JOER.99.6.323-338>
- 743 Smith, S.M., 2002. Fast robust automated brain extraction. *Human Brain Mapping* 17, 143–
744 155. <https://doi.org/10.1002/hbm.10062>
- 745 Spearman, C., 1904. “General Intelligence,” Objectively Determined and Measured. *The*
746 *American Journal of Psychology* 15, 201–292.
- 747 Tamnes, C.K., Østby, Y., Walhovd, K.B., Westlye, L.T., Due-Tønnessen, P., Fjell, A.M.,
748 2010. Intellectual abilities and white matter microstructure in development: A diffusion
749 tensor imaging study. *Human Brain Mapping* 31, 1609–1625.
750 <https://doi.org/10.1002/hbm.20962>
- 751 Tideman, E., Gustafsson, J.-E., 2004. Age-related differentiation of cognitive abilities in ages
752 3–7. *Personality and Individual Differences* 36, 1965–1974.
753 <https://doi.org/10.1016/j.paid.2003.09.004>
- 754 Tu, Y.-K., Gunnell, D., Gilthorpe, M.S., 2008. Simpson’s Paradox, Lord’s Paradox, and
755 Suppression Effects are the same phenomenon – the reversal paradox. *Emerg*
756 *Themes Epidemiol* 5, 2. <https://doi.org/10.1186/1742-7622-5-2>
- 757 Urger, S.E., De Bellis, M.D., Hooper, S.R., Woolley, D.P., Chen, S.D., Provenzale, J., 2015.
758 The Superior Longitudinal Fasciculus in Typically Developing Children and
759 Adolescents: Diffusion Tensor Imaging and Neuropsychological Correlates. *Journal*
760 *of Child Neurology* 30, 9–20. <https://doi.org/10.1177/0883073813520503>
- 761 Volkow, N.D., Koob, G.F., Croyle, R.T., Bianchi, D.W., Gordon, J.A., Koroshetz, W.J., Pérez-
762 Stable, E.J., Riley, W.T., Bloch, M.H., Conway, K., Deeds, B.G., Dowling, G.J.,
763 Grant, S., Howlett, K.D., Matochik, J.A., Morgan, G.D., Murray, M.M., Noronha, A.,
764 Spong, C.Y., Wargo, E.M., Warren, K.R., Weiss, S.R.B., 2018. The conception of the
765 ABCD study: From substance use to a broad NIH collaboration. *Developmental*
766 *Cognitive Neuroscience* 32, 4–7. <https://doi.org/10.1016/j.dcn.2017.10.002>
- 767 Vollmar, C., O’Muircheartaigh, J., Barker, G.J., Symms, M.R., Thompson, P., Kumari, V.,
768 Duncan, J.S., Richardson, M.P., Koeppe, M.J., 2010. Identical, but not the same:
769 Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T
770 scanners. *NeuroImage* 51, 1384–1394.
771 <https://doi.org/10.1016/j.neuroimage.2010.03.046>
- 772 Vollmer, B., Lundequist, A., Mårtensson, G., Nagy, Z., Lagercrantz, H., Smedler, A.-C.,
773 Forsberg, H., 2017. Correlation between white matter microstructure and executive
774 functions suggests early developmental influence on long fibre tracts in preterm born
775 adolescents. *PLOS ONE* 12, e0178893.
776 <https://doi.org/10.1371/journal.pone.0178893>

- 777 Wandell, B.A., 2016. Clarifying Human White Matter. *Annual Review of Neuroscience* 39,
778 103–128. <https://doi.org/10.1146/annurev-neuro-070815-013815>
779 Wechsler, D., 2011. Wechsler Abbreviated Scales of Intelligence-Second Edition (WASI-II).
780 Wechsler, D., 2005. Wechsler Individual Achievement Test-Second UK Edition (WIAT-II).
781 Wechsler, D., 1999. Wechsler Abbreviated Scales of Intelligence.
782 Westerhausen, R., Friesen, C.-M., Rohani, D.A., Krogsrud, S.K., Tamnes, C.K., Skranes,
783 J.S., Håberg, A.K., Fjell, A.M., Walhovd, K.B., 2018. The corpus callosum as
784 anatomical marker of intelligence? A critical examination in a large-scale
785 developmental study. *Brain Structure and Function* 223, 285–296.
786 <https://doi.org/10.1007/s00429-017-1493-0>

787

788

Supplementary Material

789

Is the Peabody Picture Vocabulary Test a measure of fluid ability?

790

791

792 As a non-preregistered exploratory analysis, we more closely examined the cross-
793 loading of the Peabody Picture Vocabulary Test (PPVT). This task asks participants to select
794 the correct picture (out of four multiple-choice options) corresponding to the meaning of a
795 word spoken by an examiner (Dunn and Dunn, 2007). As discussed previously in the
796 Results section 3.1, modification indices suggested the PPVT should either be cross-loaded
797 or solely loaded onto gf. To better understand this cross-loading, we performed an
798 exploratory (i.e. not part of preregistration) analysis using SEM tree analysis. In this analysis,
799 we allowed the PPVT to load on both gc and gf, and examined whether using age as a
800 covariate yielded a developmental period where the associations between the latent factors
801 and the PPVT task differed. This generated an age split for gf at around age 9.5 whereby the
802 loading of the PPVT decreased (from 1 to .87, unstandardized estimate).

803 Conversely, for gc the loading remained the same (.12, unstandardized estimate).

804 This suggested the PPVT as commonly implemented behaved as a fluid, rather than a
805 crystallized, task, especially in younger participants of lower ability. Although purportedly a
806 test of crystallized knowledge, the implementation of the PPVT may very well rely on more
807 fluid, executive components including response selection and reasoning, especially in a
808 cohort of children and adolescents with comparatively low overall performance.

809 A likely explanation for this pattern is that, while PPVT draws on gc, the demanding
810 nature of the task may require more fluid, executive components in younger children,
811 especially in a cohort with comparatively low overall performance (e.g. CALM). Moreover,
812 the surprisingly strong (.83, standardized) association between gf and PPVT in the full
813 sample is similar to previous research in children (Naglieri, 1981) and adults (Bell et al.,
814 2001), although with small, typically developing samples using different statistical methods.

815

References

816

817

- 818 Bell, N.L., Lassiter, K.S., Matthews, T.D., Hutchinson, M.B., 2001. Comparison of the
819 Peabody Picture Vocabulary Test—Third Edition and Wechsler Adult Intelligence
820 Scale—Third Edition with university students. *Journal of Clinical Psychology* 57, 417–
821 422. <https://doi.org/10.1002/jclp.1024>
822 Dunn, L.M., Dunn, D.M., 2007. PPVT-4: Peabody picture vocabulary test.
823 Naglieri, J.A., 1981. Concurrent validity of the revised Peabody Picture Vocabulary Test.
824 *Psychology in the Schools* 18, 286–289. [https://doi.org/10.1002/1520-
825 6807\(198107\)18:3<286::AID-PITS2310180306>3.0.CO;2-1](https://doi.org/10.1002/1520-6807(198107)18:3<286::AID-PITS2310180306>3.0.CO;2-1)
826