NeuroImage 137 (2016) 201–211

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Dissociable brain biomarkers of fluid intelligence

Erick J. Paul [a,*], Ryan J. Larsen [a], Aki Nikolaidis [a,b], Nathan Ward [a,c], Charles H. Hillman [a,d,e,f,g], Neal J. Cohen [a,e,f], Arthur F. Kramer [a,d,e,f], Aron K. Barbey [a,e,f,g,h,i,*]

[a] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[b] Center for the Developing Brain, Child Mind Institute, New York, NY, USA
[c] Tufts University, Department of Psychology, Medford, MA, USA
[d] Department of Kinesiology and Community Health, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[e] Department of Psychology, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[f] Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[g] Department of Internal Medicine, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[h] Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[i] Carle R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

ARTICLE INFO

Article history:
Received 23 September 2015
Revised 6 May 2016
Accepted 11 May 2016
Available online 14 May 2016

Keywords:
N-Acetyl aspartate
Fluid intelligence
Brain volume
MR spectroscopy

ABSTRACT

Cognitive neuroscience has long sought to understand the biological foundations of human intelligence. Decades of research have revealed that general intelligence is correlated with two brain-based biomarkers: the concentration of the brain biochemical N-acetyl aspartate (NAA) measured by proton magnetic resonance spectroscopy (MRS) and total brain volume measured using structural MR imaging (MRI). However, the relative contribution of these biomarkers in predicting performance on core facets of human intelligence remains to be well characterized. In the present study, we sought to elucidate the role of NAA and brain volume in predicting fluid intelligence (Gf). Three canonical tests of Gf (BOMAT, Number Series, and Letter Sets) and three working memory tasks (Reading, Rotation, and Symmetry span tasks) were administered to a large sample of healthy adults ($n = 211$). We conducted exploratory factor analysis to investigate the factor structure underlying Gf independent from working memory and observed two Gf components (verbal/spatial and quantitative reasoning) and one working memory component. Our findings revealed a dissociation between two brain biomarkers of Gf (controlling for age and sex): NAA concentration correlated with verbal/spatial reasoning, whereas brain volume correlated with quantitative reasoning and working memory. A follow-up analysis revealed that this pattern of findings is observed for males and females when analyzed separately. Our results provide novel evidence that distinct brain biomarkers are associated with specific facets of human intelligence, demonstrating that NAA and brain volume are independent predictors of verbal/spatial and quantitative facets of Gf.

Introduction

Research in the psychological and brain sciences has long sought to elucidate the nature and mechanisms of human intelligence. Early research by Charles Spearman (1904) provided the foundation for this endeavor, revealing that individuals' performance across a broad range of cognitive tasks is positively correlated. This observation led Spearman (1927) to propose that a general factor (g) accounts for performance across the spectrum of cognitive ability — spanning attention, perception, memory, language, and thought. Decades of research have further demonstrated that the best measures of g involve tests of fluid intelligence (Gf) — the capacity to solve novel problems through adaptive reasoning and goal-directed decision making (Carroll, 1993; Cattell, 1971; Gray and Thompson, 2004; Horn and Cattell, 1966; Jensen, 1980). Performance on tests of Gf is known to predict many aspects of life, including educational and work achievement, and social well-being (Colom and Flores-Mendoza, 2007; Gottfredson and Saklofske, 2009; Jensen, 1998; Neisser et al., 1996).

Parallel developments in cognitive neuroscience have advanced our understanding of the neurobiological foundations of Gf (Barbey et al., 2014; Barbey et al., 2013a, 2013b; Jung and Haier, 2007). An emerging area of research investigates the metabolic and biochemical correlates of intelligence using magnetic resonance spectroscopy (MRS). Accumulating evidence indicates that a specific biochemical marker is associated with general intelligence: N-acetyl aspartic acid

* Corresponding authors at: Decision Neuroscience Laboratory, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 405 North Mathews Avenue, Urbana, IL 61801, USA.
E-mail addresses: ejpaul@illinois.edu (E.J. Paul), barbey@illinois.edu (A.K. Barbey).
URL's: http://DecisionNeuroscienceLab.org/ (E.J. Paul), http://DecisionNeuroscienceLab.org/ (A.K. Barbey).

http://dx.doi.org/10.1016/j.neuroimage.2016.05.037

(NAA) (Jung and Haier, 2007; Ross and Sachdev, 2004). NAA is a metabolite of glucose that is produced in the neurons and represents an important biochemical marker of energy production and neuronal health (Barker et al., 2001; Nakashima et al., 2007). Several studies have reported correlations between the concentration of NAA in the brain and various domains of cognition and intelligence (Jung and Haier, 2007; Ross and Sachdev, 2004), and the emerging evidence frequently favors a positive relationship between NAA and cognition.

Nevertheless, this literature exhibits some variability and inconsistency (Patel et al., 2014). Positive correlations have been reported between Gf and NAA in occipital–parietal white matter (WM) (Jung et al., 1999; Jung et al., 2005), the isthmus/splenium region of the corpus callosum (Aydin et al., 2012), and deep cerebral WM (Charlton et al., 2007). In contrast, NAA in frontal WM has been found to negatively correlate with Gf (Jung et al., 2005), while NAA in parietal WM has been found to be uncorrelated with Gf (Ferguson et al., 2002). Another study reported a positive correlation between a principle component loading heavily on Gf and frontal WM NAA, but not with occipital–parietal gray matter (GM) NAA (Ross et al., 2005). A multi-voxel study using several tests of Gf found that NAA in right posterior GM correlated with some — but not all — tests of Gf, whereas no relationship with Gf and WM NAA in any region was found (Jung et al., 2009). Another recent multi-voxel study found that NAA in left frontal and parietal regions is related to Gf (Nikolaidis et al., 2016).

One possible reason for disagreement across studies is that the strength of NAA-intelligence correlations may depend upon the particular domain of intelligence measured. This would likely cause the NAA-intelligence correlation to be contingent upon the particular test(s) employed in a given study. Experiments relating Gf and NAA have generally used performance factors from the WAIS, or the Raven's progressive matrices (RPM). The RPM test is considered to be one of the most fundamental and accurate tests of fluid reasoning ability (Jensen, 1998) and draws upon spatial reasoning ability (Ackerman et al., 2002). Reasoning ability can also be manifest through quantitative or numerical reasoning (Ackerman et al., 2002); however, little is known whether these abilities relate to NAA concentration.

By characterizing sub-domains of fluid intelligence, researchers have revealed sex differences in quantitative and verbal reasoning abilities (Halpern, 2013), and have demonstrated significant sex differences in the neurocorrelates of intelligence (Burgaleta et al., 2012; Schmithorst, 2009; Witelson et al., 2006). It is therefore reasonable to expect that examination of sex differences may be important for understanding the Gf–NAA relationship (Pfleiderer et al., 2004). However, studies reporting sex differences also exhibit inconsistent findings. Two studies have found positive Gf–NAA correlations for females but not males (Pfleiderer et al., 2004; Jung et al., 2005), but a study with a larger sample size reported similar Gf–NAA correlations for males and females (Jung et al., 2009).

Disagreement across studies may also be due in part to interrelated cognitive factors, or cognitive sub-components that drive performance on tests of intelligence. For example, there is a long history of behavioral evidence suggesting that Gf is closely related to working memory capacity (Kane et al., 2005; Kyllonen & Christal, 1990; Martinez et al., 2011). More recent neuroimaging and cognitive neuroscience evidence demonstrates a high degree of correspondence between the brain structures supporting working memory and Gf (Barbey et al., 2014; Gray et al., 2003; Kane & Engle, 2002), bolstering the behavioral evidence and further indicating these two constructs are linked. Therefore, evaluating the relationships between multiple cognitive domains to NAA within a sample may be a key factor for understanding MRS–intelligence relationships, and whether or not NAA is a specific marker for intelligence, or an array of cognitive abilities.

Results across studies may also prove to be more consistent by reducing variability in methodology. First, improved quantification can be achieved by applying corrections for tissue fractions (Gussew et al., 2012); however, this is not performed in all studies (Patel et al., 2014). Second, targeting regions of the brain that can be reliably positioned and scanned using MRS is critical, as it is possible to obtain intra-subject coefficients of variation as low as 5% for the most easily measured metabolites (Brooks et al., 1999; Terpstra et al., 2015). However, selecting brain regions that are related to the cognitive domains of interest must be balanced with this requirement. The precuneus and posterior cingulate cortex (PCC) are excellent locations for high quality spectra, and have been used extensively in previous MRS studies. The PCC is involved in a wide range of cognitive functions, from internal awareness, to attention regulation. This region has strong reciprocal connections to the ACC and DLPFC, which are critical for executive function and fluid reasoning. It also has an exceptionally high basal metabolic rate, and given that PCC connectivity and activity specifically declines with age (Andrews-Hanna et al., 2007; Leech and Sharp, 2014) and several disease states such as Alzheimer's disease (Minoshima et al., 1997), traumatic brain injury (Nakashima et al., 2007), and schizophrenia (Haznedar et al., 2004), NAA in the PCC may be a good marker for overall brain metabolic health that is readily measurable via MRS.

Another obstacle to obtaining reproducible results across studies is small sample sizes (Patel et al., 2014). Typical samples range between 30 and 80 subjects, with the largest including 88 subjects (Ferguson et al., 2002). Results obtained with these small sample sizes are difficult to interpret — especially after dividing the sample into smaller subsets based on individual subject characteristics, such as sex (Jung et al., 2009). More generally, correlation coefficients tend to lack stability in samples smaller than about 100 subjects when effect sizes are small or moderately sized (Schönbrodt and Perugini, 2013). Given that the relationships between Gf and NAA are small-moderate in effect, establishing solid relationships between Gf and NAA requires the use of much larger sample sizes.

The use of larger sample sizes together with collection of a wide range of both MRI and behavioral data should also allow researchers to control for other factors related to intelligence, thereby elucidating the unique contribution of NAA to Gf, and disentangling the Gf–NAA empirical landscape from possible confounding variables. One of the most well-established neurocorrelates of intelligence is total brain volume (Ivanovic et al., 2004; Ritchie et al., 2015; for reviews, see Luders et al., 2009; McDaniel, 2005; Rushton and Ankney, 2009). This relationship is likely due to an increased number of neurons (Rushton and Ankney, 2009) or more efficient neuronal metabolism in larger brains (Gignac et al., 2003; Haier et al., 1995). Because NAA is a measure of neuronal density, viability, or efficiency, it is reasonable to expect that NAA–intelligence correlations may not be independent of brain size. A positive correlation between NAA and total WM volume has also been previously reported (Jung et al., 2005); however, two other studies showed that the NAA–intelligence relationship persists when brain size is covaried (Aydin et al., 2012; Nikolaidis et al., 2016). Further investigation is needed to understand whether NAA and brain volume independently account for variance in intelligence scores, or if their predictive power overlaps.

The present study seeks to characterize the roles of NAA concentration and brain volume in predicting Gf, and whether or not these two markers are independent predictors of Gf. Our study investigated a sample size of over 200 participants, more than doubling the largest sample size reported in the literature to date. Additionally, we measured both Gf and working memory using three unique, well-established tests in order to investigate how NAA and brain volume relate to underlying sub-factors of fluid reasoning, and whether those relationships depend upon working memory. Our estimates of brain volume include total brain volume as well as tissue specific, total gray and total white matter volumes. Our measurements of NAA are obtained from a single voxel in medial parietal cortex extending inferiorly into the posterior cingulate cortex — an excellent region for obtaining high quality spectra that has been widely used and validated in prior research. Furthermore, our analysis of the relationship between Gf and NAA takes into account

age and sex for the assessment of Gf, as well as tissue composition in estimating NAA concentration.

Method

Participants

Participants were recruited from the Urbana-Champaign community as part of a larger cognitive training intervention study designed to assess the efficacy of different intervention modalities on cognitive performance in healthy adults. All data reported here were collected as part of the baseline, pre-intervention assessment, which included a battery of cognitive tests and an MRI session. The University of Illinois Urbana-Champaign Institutional Review Board approved all aspects of the study and participants provided informed consent at enrollment. A total of 225 participants from that sample underwent an MRI scanning session including MRS during baseline (i.e., pre-intervention) testing; of those, 211 participants had complete MRS (7 excluded due to MRI/MRS quality), behavioral (one excluded due to missing), and demographic (six excluded due to missing) data. All participants: were right-handed with normal or corrected-to-normal vision without color blindness; reported no previous neurological injuries, disorders, or surgeries; were on no medications affecting central nervous function; were not pregnant; had no head injury or loss of consciousness in the past 2 years; and were proficient in English. Participants received monetary compensation for their participation. Basic demographics were collected via self-report and are summarized in Table 1. There was no significant difference between males and females in age ($t(209) = 0.57$, $p = 0.57$). Although a higher proportion of females were college graduates than males (Table 1), approximately 75% of males and females reported having some college or graduated college; therefore, males and females were roughly equally educated.

Cognitive testing procedures

All cognitive tests were computer administered in a quiet room. Test administrators provided general instruction and oversight during the approximately three hour testing session and test-specific instructions were provided through on-screen prompts. Participants wore headphones during the session and made responses with the mouse or keyboard.

The pre-intervention cognitive battery included a total of 12 tests: three unique tests for each cognitive construct including Gf, working memory, executive function, and episodic memory. Gf and working memory are included in all analyses herein. (Although performance on executive function and episodic memory is not within the scope of the present manuscript, interested readers may find analyses including these tasks in supplementary materials.)

Fluid intelligence

Because fluid reasoning can manifest in a broad array of abilities including spatial, quantitative, and verbal (Ackerman et al., 2002), our tests were selected to tap into each. The three Gf tests included the BOMAT, Number Series, and Letter Sets.

Table 1
Sample demographics.

	Male	Female	Total
n	121	90	211
Mean age (range)	24.07 (18–44)	24.57 (18–44)	24.28 (18–44)
Highest level of education (proportion of sample)			
High school graduate	0.08	0.03	0.06
Some college	0.6	0.44	0.53
College graduate	0.15	0.31	0.22
Master's or higher	0.17	0.22	0.19

BOMAT. In this task (Hossiep et al., 1999; Jaeggi et al., 2008; Moody, 2009), which is similar to RPM but has greater difficulty levels to avoid ceiling effects, participants were presented a series of 5×3 matrices that each depict a pattern (29 different matrices in total with 45 min to complete the test), with each matrix missing one cell. The participant's task was to select one of six possible answers that complete the matrix pattern.

Number Series. In the Number Series task (Harrison et al., 2013; Thurstone, 1938), participants were shown a series of arithmetic number patterns that follow an arithmetic sequence, and their task was to select the next number in the series from five possible answer options (10 trials in total with 5 min granted to complete the test).

Letter Sets. In the Letter Sets task (Ekstrom et al., 1976; Harrison et al., 2013), participants were shown five sets of letters with four letters in each set. Four of the five letter sets followed a common rule, and the participants' task was to select the letter set that is different from the other four (15 trials in total with 7 min to complete the test).

Working memory

The working memory tests included reading span, rotation span, and symmetry span. For all three tasks, shortened versions were administered, which have been verified to retain the psychometric properties of the longer versions (Foster et al., 2015; Oswald et al., 2015).

Reading span. During the first practice phase, participants were shown a series of simple four letter words (nouns/verbs) each for 1 s, after which they were asked to recall which ones they saw in the correct order from a selection of fifteen words. In the next practice phase, participants were shown a series of short sentences and were tasked to select whether or not they are understandable or nonsensical (e.g., "We were two lawns out at sea before we lost sight of land.") The nonsensical sentences portion always came from changing one word (e.g., lawns for miles) at the beginning, middle, or end of the sentence. Roughly half of the presented sentences were nonsensical. Finally, during test trials, participants would characterize the sentences and were immediately shown a four-letter word for 1 s. After going through the entire set (set size range from 2 to 10), participants were given a recall cue and needed to identify the words they saw in correct order. Participants were told to maintain accuracy of characterizing the sentences correctly at 85% or higher. The time limit for evaluating sentences was individualized and determined by practice phase performance. There were 8 test trials in total.

Rotation span. Participants were asked to recall a sequence of short and long arrows radiating from the center of the screen against a background letter-rotation task. The letter-rotation task presented a normal or mirror-reversed G, F, or R, rotated at 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°. The task was to mentally rotate the letter, and then to indicate whether the letter was normal (True — approximately 50% of trials) or mirror reversed (False — also approximately 50% of trials). Immediately after a response, the participant pressed a key clearing the screen for 0.5 s and was presented a short or long arrow rotated at 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°. After 1 s, the arrow disappeared and another letter or the recall cue appeared instructing the participant to recall all of the arrows from the preceding displays in the order they appeared. Participants were told to maintain accuracy on the letter-rotation task at 85% or higher. The time limit for the rotated letter judgment was individualized and determined by practice phase performance. Set sizes ranged from two to nine letter/arrow displays per trial (8 trials total).

Symmetry span. On each trial a 4×4 grid was presented in which one of the 16 possible locations was filled in red (650 ms each for 3–6 locations). Participants were asked to remember the location of the red

squares. Between each location presentation, participants were shown an 8 × 8 grid of black and white rectangles. They were asked to determine whether or not the grid was symmetric about the vertical axis (i.e., left half matches right). After all spatial locations were presented, participants were asked to reproduce the spatial locations in the order in which they were presented. Participants are told to maintain accuracy of identifying symmetry at 85% or higher and were shown their accuracy at the end of each trial. The time limit for the symmetry judgment was individualized and determined by practice phase performance. There were a total of 8 trials.

MRI data acquisition and processing

Acquisition

All subjects were scanned on a Siemens 3T Magnetom Trio. Anatomical information was obtained using a high resolution 3D structural MPRAGE scan (0.9 mm isotropic, TR: 1900 ms, TI: 900 ms, TE = 2.32 ms, with GRAPPA and an acceleration factor of 2).

The anatomical scan was used to position a single voxel spectroscopy (SVS) scan in the parietal cortex extending into posterior cingulate cortex (voxel size: $(20\ mm)^3$, TR: 3000 ms, TE: 30 ms, 40 averages, BW: 2000 Hz, vector size: 1024). The voxel straddled the midline, including equal portions of the right and left hemispheres as shown in Fig. 1. Weak water suppression was employed and six regional saturation bands were placed around the voxel to reduce contamination from subcutaneous fat. An additional scan was performed without water suppression to aid with quantification. Immediately following the spectroscopy acquisition, a T_2-weighted overlay scan was performed with the same center location and orientation as the spectroscopy scan (TR = 5000 ms, TE = 84 ms, slice thickness 2 mm with 0.5 mm of spacing, FOV: 240 × 240 mm, 128 × 128 matrix size, GRAPPA acceleration factor: 2, 35 slices).

MRI data processing

Metabolite quantitation was performed using tissue water as a reference (Gasparovic et al., 2006). This approach is commonly used in studies that relate metabolite concentration to intelligence (Charlton et al., 2007; Jung et al., 1999; Jung et al., 2009; Jung et al., 2005; Kochunov et al., 2010; Ross et al., 2005). Metabolite ratios have also been used in similar studies; however, use of water scaling facilitates the interpretation of results and the separation of contributions from the different metabolites.

Water-scaled spectra were analyzed using LCModel software (Version 6.3-1H; Provencher, 1993). No correction was performed to account for relaxation of metabolite signal. Because NAA and NAAG are difficult to differentiate (Edden et al., 2007), here we analyze the combined concentration of NAA + NAAG, labeled herein as NAAt with a peak appearing at 2.02 ppm.

Accurate water scaling requires corrections for the tissue composition of the voxel. Using the high resolution structural scan, we calculated the volume fractions of gray matter (GM), white matter (WM), and cerebral spinal fluid (CSF) within each voxel using Matlab scripts (MathWorks, Natick, MA) that called functions from SPM8 (Wellcome Trust Centre for Neuroimaging). First, we segmented the MPRAGE using SPM8 to obtain tissue probability maps of GM, WM and CSF. We then created a mask in the space of the T_2 overlay corresponding to the location of the spectroscopy scan. This mask has the same center and orientation as the T_2-overlay but higher resolution (0.5 × 0.5 × 0.5 mm). We then registered the MPRAGE to the T_2 overlay. The rotations and translations required for the registration were then applied to the tissue probability masks. We resliced the tissue probability maps into space of the mask, and used the mask to calculate the volume fractions of GM, WM, and CSF within the volume of the spectroscopy voxel (see the white box shown in Fig. 1). These tissue fractions were later used to statistically correct

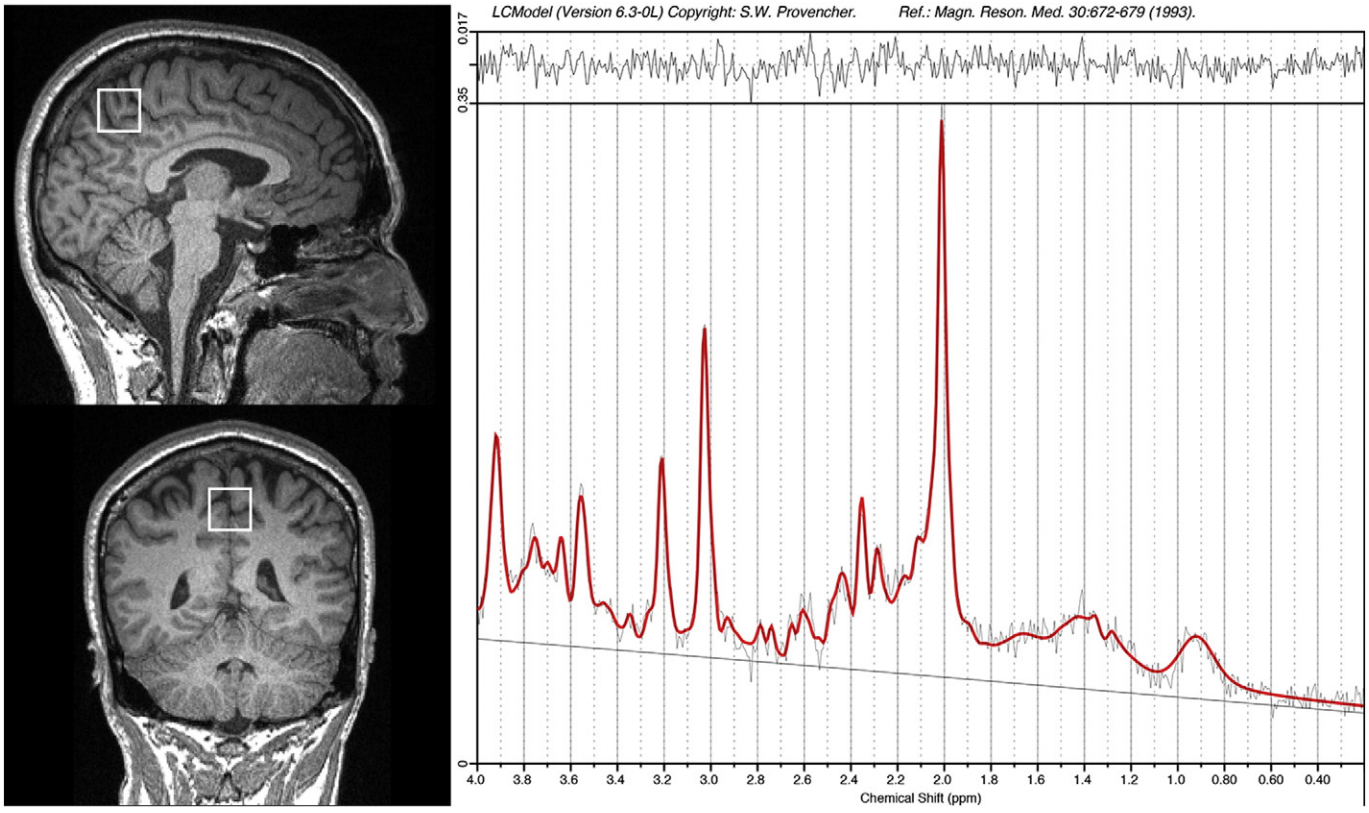

Fig. 1. Left: example placement of MRS voxel spanning precuneus and extending into posterior cingulate (bilaterally) for a single subject's sagittal (top) and coronal (bottom) views. Right: example MRS spectrum output from LCModel (thin black line is original data; thick overlaid line is LCModel data).

NAAt for tissue volume-fraction dependencies (see MRS metabolite correction section).

Total brain volume was estimated from the MPRAGE image using the FSL FAST segmentation tool (Jenkinson et al., 2012; Zhang et al., 2001). This approach yields probabilistic segmented images with values ranging from 0 to 1 in every voxel for GM, WM, and CSF. Final tissue assignment in each voxel was accomplished by selecting the maximum probability value of the three tissue types, guaranteeing that only one tissue is assigned to each voxel uniquely. Total GM and WM volumes were calculated separately by summing the volume across voxels assigned each tissue type, respectively. Total brain volume was taken as the sum of the GM and WM volumes.

Statistical analysis of behavioral and MRS data

All statistical analyses were carried out using R version 3.1.1 (www.r-project.org).

Fluid intelligence composite and factor score construction

Behavioral performance on the Gf tests and working memory tests was analyzed in two ways. First, each of the tests of Gf and working memory were standardized before being summed to create a composite Gf and composite working memory score (Baniqued et al., 2014). Second, all six Gf and working memory tests were submitted together to an exploratory factor analysis to characterize sub-domains of Gf and working memory (via PCA; Abdi and Williams, 2010; Jolliffe, 2002). Use of an exploratory factor analysis allows us to investigate the interrelatedness of the tests, the cognitive factors, and to probe sub-domains of Gf in greater detail. Factor scores for each retained component were calculated after applying a rotation (both oblique and orthogonal rotations were tested — see PCA of Gf and working memory tests section). We compared results from the factor analysis to those of the standardized composite Gf score.

MRS metabolite correction

Subjects with spectroscopy data were excluded if the full width at half maximum (FWHM) reported by LCModel was greater than 10 Hz (an effective cutoff for identifying subjects with poor spectral quality; five subjects excluded) or the segmentations in SPM8 failed (two subjects excluded).

Calculated metabolite concentrations depend on the volume fractions of brain tissues within the single voxel spectroscopy scan (e.g., NAAt is known to be greater in GM than in WM (Gasparovic et al., 2006; Pfefferbaum et al., 1999). We observed highly significant (two-tailed) correlations between NAAt and GM [$r(209) = 0.66$, $p < 0.001$] and CSF [$r(209) = -0.79$, $p < 0.001$] volume fractions within the SVS volume. Because of this observed systematic dependence, we corrected the water-scaled concentration values for tissue fraction by implementing a statistical correction. We fit a multiple regression model predicting the metabolite NAAt by GM and CSF fractions, and then computed the residuals from that model to use as our metabolite corrected for tissue fractions. This procedure de-correlates the metabolite with respect to the tissue fractions (all p-values >0.99), and controls for the possibility that relationships observed between NAAt and cognitive variables are simply a byproduct of differences in tissue volume fractions. Concentrations of NAAt measured with this technique are referred to as NAAt-corrected.

Models of interest

The relationship between cognitive factors and metabolites was first assessed using Pearson's correlation, and then tested further via multiple linear regression including other covariates. Performance on Gf tests tasks may vary with age (Horn and Cattell, 1967) and sex (Irwing and Lynn, 2005); thus, including these as covariates in a more detailed analysis provides more information about the nature of the NAA–intelligence relationship. The regression model predicted cognitive factor scores (dependent variable) from the metabolite concentration, estimated brain volume (independent variables) and age and sex (covariates). Finally, we re-computed all statistical tests for males and females separately to determine if the magnitude of correlations differs across sexes.

Results

PCA of Gf and working memory tests

To perform dimensionality reduction with PCA, one common approach is to retain enough components to explain a fixed amount of the total variance, typically in the range of 70% to 95% (Jolliffe, 2002). We conservatively set our a priori inclusion criterion at 70% of the variance explained (i.e., our factor solution should explain a minimum of 70% of the variance in the data). Initially we performed an oblique rotation because this solution does not impose orthogonality on the recovered components, and previous work suggests that Gf and working memory are correlated. To meet our criterion, three components were retained accounting for 75% of the variance in the data. The largest correlation between factors was 0.344 (between components one and two), which is just above the 0.32 cut off recommended for determining whether factors are orthogonal (Tabachnick and Fidell, 2007). This suggests that the components are not sufficiently orthogonal to justify using an orthogonal (varimax) rotation. However, we also applied a varimax rotation (Supplementary Table 1) to further corroborate the factor structure and to help confirm our interpretation of the Gf and working memory factors.

Descriptive statistics for each test appear in Table 2, and Table 3 shows the rotated pattern matrix for the three retained components. Note that this pattern matrix was qualitatively identical using either an oblique or orthogonal rotation scheme (see Supplementary Table 1 for the orthogonal rotation pattern matrix). We observed a clear separation between the working memory and Gf tests. All three working memory tests largely load onto only the first component whereas the Gf tests split between the second and third components. Number Series and Letter Sets clearly load onto different Gf factors (see Table 3); BOMAT loads somewhat on both Gf factors, though it most strongly contributes to the second component. For clarity and ease of understanding — together with the observed numerical loadings — we refer to the first factor as working memory, the second as verbal/spatial reasoning and the third as quantitative reasoning. Our factor structure implies that the two Gf factors we observe are both separable from working memory; such a separation between Gf and working memory has been previously observed in a factor analysis that used tests similar to those used here (Foster et al., 2015).

Both Gf factor scores are positively correlated with the composite Gf score: verbal/spatial factor, $r(209) = 0.823$, $p < 0.001$ or 67.8% shared variance; quantitative factor, $r(209) = 0.693$, $p < 0.001$, or 48.0% shared variance. The working memory factor score is very highly correlated with the working memory composite score: $r(209) = 0.998$, $p < 0.001$.

Bivariate correlations between Gf and brain biomarkers

We observed that NAAt-corrected is positively correlated with verbal/spatial reasoning (Fig. 2), but not correlated with quantitative reasoning

Table 2
Descriptive test statistics reporting mean (SD) for each.

	All subjects	Males	Females
BOMAT	14.99 (4.22)	15.02 (4.3)	14.97 (4.19)
Number Series	6.93 (2.04)	7.32 (2.01)	6.41 (1.99)
Letter Sets	10.80 (2.27)	10.69 (2.35)	10.95 (2.16)
Reading span	18.77 (9.33)	18.41 (9.69)	19.26 (8.85)
Rotation span	12.67 (6.74)	13.17 (6.79)	12.00 (6.66)
Symmetry span	18.49 (7.87)	19.19 (8.19)	17.54 (7.36)

Table 3
Rotated pattern matrix (oblique rotation).

	Working memory	Verbal/spatial	Quantitative
BOMAT	0.05	**0.66**	0.38
Number Series	0.02	−0.01	**0.95**
Letter Sets	0.01	**0.92**	−0.1
Reading span	**0.71**	0.19	−0.15
Rotation span	**0.89**	−0.11	0.03
Symmetry span	**0.78**	0.03	0.08
Proportion variance explained	0.33	0.23	0.19

Note: Numbers in bold correspond to variables with largest loadings and therefore principally associated with each pattern.

(see Table 4).[1] In contrast, brain volume is strongly correlated with quantitative, but not verbal/spatial reasoning. These results therefore reveal a striking dissociation: brain volume is a marker of the quantitative reasoning factor, and NAAt-corrected is a marker of the verbal/spatial reasoning factor. The working memory factor largely reflects the results of the quantitative reasoning factor: it is positively correlated with total brain volume, GM volume, and WM volume, but not NAAt-corrected.

Furthermore, the correlations between the Gf factors and biomarkers are substantially stronger than the correlations between the composite Gf score and biomarkers (see Table 4). This underscores the power of the factor analytic approach employed here: the observed specificity of these empirically derived factors of fluid intelligence for very different brain biomarkers is lost when using a simple composite score. Finally, the observed relationship between brain volume and the quantitative factor score is not overwhelmingly driven by a single tissue type as both GM and WM volumes are significantly correlated with the quantitative factor score. Neither GM volume nor WM volume were significantly correlated with the verbal/spatial factor score.

Note that three subjects' original NAAt values fall outside 3 standard deviations from the mean NAAt value. We therefore removed these three data points and re-calculated all bivariate correlations. The pattern of results and significance were qualitatively unchanged; thus, all subjects are included here and in all subsequent analyses.

Confirmatory modeling: cognitive factors and brain biomarker models controlling for covariates

To assess the robustness of each Gf–biomarker relationship, we fit a regression model that simultaneously included both biomarkers, NAAt-corrected and brain volume, while also including age and sex as covariates. To ensure our measured biomarkers are not systematically related to each other and are independently predicting variance in the full multiple regression model, we first computed the correlations between NAAt-corrected and brain volume. No linear relationship between the biomarkers was observed ($p = 0.96$, two-tailed). Similarly no relationship was observed between NAAt-corrected and total white matter volume ($p = .41$, two-tailed) or total gray matter volume ($p = 0.64$, two-tailed); because GM and WM volume correlations were similar to total brain volume (see Table 4), and total brain volume was highly correlated with both GM and WM (both $r > 0.97$), only total brain volume was included in the regression analyses.

This model was fit separately for each of the cognitive factor scores as well as the cognitive composite scores; the standardized regression parameter estimates and p-values for each biomarker appear in Table 5. After controlling for covariates, the composite Gf z-score is significantly correlated with NAAt-corrected, and has a trending positive correlation with brain volume. However, the dissociation between NAAt-corrected and brain volume with the verbal/spatial component and the quantitative component persists, suggesting that both NAAt and brain volume independently correlate with different facets of Gf even after controlling for performance differences driven by age, sex, and education. Brain volume remained the only biomarker significantly related to the working memory composite and factor scores.

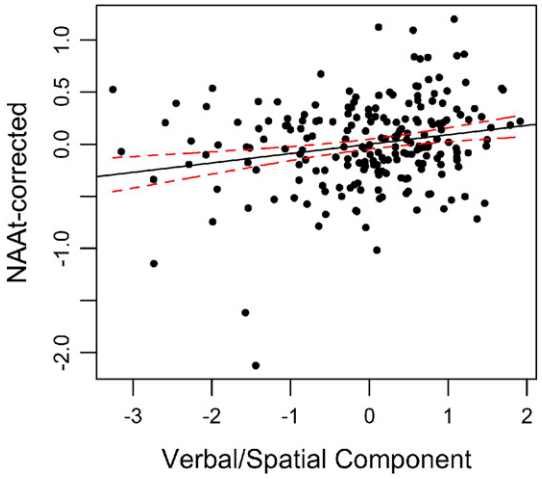

Fig. 2. Scatter plot of NAAt-corrected and the verbal/spatial Gf component. Dashed lines represent 90% confidence interval of prediction.

For conciseness, the covariates' parameter estimates are not presented in Table 5. However, the following covariates were trending or significant predictors in the models: age had a trending negative relationship with working memory ($p = 0.076$), and a significant negative relationship with quantitative reasoning ($p = 0.03$); similarly sex had a trending relationship with quantitative reasoning ($p = 0.084$).

Sex differences in NAA and Gf

Descriptive statistics and PCA of Gf tests

Descriptive statistics for both males and females on each Gf test are presented in Table 2. Males and females performed equivalently on BOMAT and Letter Sets; however, males performed significantly better than females on Number Series in the present sample, $t(209) = 3.28$, $p = 0.001$. Descriptive statistics for the working memory tests also appear in Table 2. There were no significant sex differences in any of the working memory tests.

Our data reveal some sex differences in the biomarkers. Males have higher brain volume, $t(209) = 8.92$, $p < 0.001$ two-tailed, whereas females exhibit higher levels of NAAt (uncorrected for tissue fractions), $t(209) = -3.6$, $p < 0.001$ two-tailed. However, no sex differences are apparent in NAAt after performing the statistical correction for tissue volume fractions ($p = 0.15$ two-tailed); thus, this difference is spurious and likely driven by male/female differences in tissue volume fractions within the MRS imaging volume.

We conducted exploratory PCA to further examine the factor structure of Gf in males versus females: as before, an oblique rotation was used and three factors were retained for both groups explaining a total of 75% and 73% of the variance, respectively (Table 6). Notably, the factor structure and loadings for males qualitatively replicates that observed for the entire sample (i.e., including both males and females); however, the factor loadings for females are slightly different. Although Number Series and Letter Sets still load most heavily on separate components, BOMAT shares its heaviest loading with Number Series for females (Table 6). Therefore, for females only, we refer to the first reasoning factor as "quantitative/spatial" and the second factor as "verbal." As in the full sample, the working memory tests loaded most heavily on a distinct and separate component for both males and females.

[1] Creatine (Cr) and Choline (Cho) are common metabolites typically estimated in the spectra together with NAAt. Correlations between all cognitive data and those metabolites, as well as ratios of those metabolites, are reported in Supplementary Table 2.

Table 4
Bivariate correlations between cognitive scores and brain biomarkers (n = 211). All p-values are two-tailed.

	NAAt-corrected	Brain volume	GM volume	WM volume
Gf-composite	0.128 (p = 0.063)	**0.15 (p = 0.026)**	**0.16 (p = 0.02)**	**0.136 (p = 0.049)**
WM-composite	0.074 (p = 0.3)	**0.204 (p = 0.003)**	**0.206 (p = 0.003)**	**0.19 (p = 0.006)**
Working mem	0.063 (p = 0.361)	**0.211 (p = 0.002)**	**0.213 (p = 0.002)**	**0.197 (p = 0.004)**
Verbal/spatial	**0.211 (p = 0.002)**	0.009 (p = 0.895)	−0.001 (p = 0.989)	0.003 (p = 0.961)
Quantitative	−0.048 (p = 0.484)	**0.237 (p < 0.001)**	**0.256 (p < 0.001)**	**0.208 (p = 0.002)**

Bivariate correlations between Gf and brain biomarkers

We investigated whether the correlations between brain biomarkers and each Gf factor score are retained within each sex. Table 7 reports the correlations for each Gf component separately for males and females. As before, the composite Gf score shows weaker correlations across both biomarkers. Males demonstrate significant correlations between NAAt-corrected and verbal/spatial as well as between brain volume and quantitative reasoning, echoing the double dissociation reported for the full sample. Although females demonstrate non-significant correlations of lesser magnitude, the pattern is still consistent with the full sample's pattern of results.

The slightly larger sample of males yielded more power to detect effects; however, the observed magnitude of correlations were not statistically different between males and females. Both the correlation between NAAt-corrected and the principally verbal component as well as the correlation between brain volume and the principally quantitative component were not significantly different between males and females ($z = 0.13$, $p = 0.897$; $z = 0.31$, $p = 0.757$, respectively). Similarly, the correlation between brain volume and the working memory component was not significantly different between males and females ($z = -0.54$, $p = 0.59$).

Discussion

To our knowledge, the present data set is the largest sample relating Gf to both NAA and brain volume. Here, we showed that NAA (measured in the posterior cingulate cortex (PCC) and parietal cortex) and brain volume are dissociable predictors of two distinct components derived from our Gf tests: NAA predicts our verbal/spatial reasoning component, whereas brain volume predicts our quantitative reasoning component. Importantly, the biomarkers are more strongly correlated with those factor components than with a composite Gf score, indicating that our use of an empirical factor analytic approach offers a more nuanced, rich view of this particular data set, suggesting that the biomarkers exhibit specificity for separate, measurable sub-domains of fluid intelligence independent of working memory. Our data further show that this pattern of results replicates for males and females.

Brain biomarkers of fluid intelligence and working memory

The positive correlation between Gf and NAA observed here is in line with many previous MRS-cognition studies (Ross and Sachdev, 2004) and is congruent with the hypothesis that NAA is a marker for general neuronal health or density (Moffett et al., 2007), but independent of brain volume, and perhaps reflective of capacity for cognitive performance. The magnitude of the Gf–NAA correlation value reported here is lower than those typically reported (see Patel et al., 2014). This is consistent with the fact that large studies tend to report smaller effect sizes (Patel et al., 2014), likely because the positive NAA–cognition relationships are somewhat over-estimated in smaller samples.

Similarly, the observed positive correlations between brain volume and intelligence in the present study are consistent with a large literature demonstrating similar results (McDaniel, 2005). The underlying factors driving the well-established brain volume-intelligence relationship are largely unknown. One possibility is that the correlation is driven by gene–environment interactions, suggesting more complex, bi-directional causality in the relationship between intelligence and brain volume (Rushton and Ankney, 2009). Another possibility is that a larger brain size is advantageous for intelligence because of fundamental architectural, histological, and/or biochemical properties of the brain. One study has reported a positive relationship between white matter volume and NAA (Jung et al., 2005), suggesting that brain biochemistry, metabolism and volume may be interrelated. However, two studies have reported positive NAA–intelligence relationships while controlling for brain size, suggesting that NAA and brain volume are independent predictors of intelligence (Aydin et al., 2012, Nikolaidis et al., 2016).

Brain volume, but not NAA, is also significantly correlated with the working memory factor in our analysis. While previous research has shown positive correlations between working memory and brain volume (Wickett et al., 2000), MRS-cognition studies with memory are relatively small in number and, as with intelligence, report disparate results. One study reported a positive correlation between NAA (as a ratio to choline) and tests of memory in the medial temporal lobes (Giménez et al., 2004) and another study reported that NAA in frontal WM is positively correlated with working memory (Yeo et al., 2000). In contrast, one multiple single-voxel MRS study reported no significant correlations with working memory in either occipital–parietal WM or frontal WM (Jung et al., 2005), and a recent multi-voxel study reported no significant correlation with working memory (Nikolaidis et al., 2016). Our data suggest that NAA is a specific marker for Gf, whereas NAA/Cr is related to working memory (Supplementary Table 2). Our data also suggest that working memory is only weakly related to Gf. Although much psychometric evidence suggests a strong relationship between Gf and working memory, there is an active debate regarding the nature and strength of this relationship (Ackerman et al., 2005; Chuderski, 2013; Colom et al., 2015; Conway et al., 2003; Kane et al., 2005) — for example, one study (Chuderski, 2013) has argued that Gf and working memory are only strongly related when time pressure is applied in Gf tests. Appealing to relationships with brain biomarkers may help to further disentangle under what psychometric conditions working memory and Gf are related. Minimally, our results hint that the relationship between verbal/spatial reasoning and NAA is not dependent on working memory, but the relationship between quantitative reasoning and brain volume could partly depend on working memory.

Our results clearly demonstrate that NAA and brain volume (including segmented GM and WM volumes) are independent, dissociable predictors of separable sub-components of fluid intelligence, suggesting that a more thorough understanding of the neural mechanisms of Gf

Table 5
Standardized regression coefficients.

	NAAt-corrected			Brain volume		
	b-Estimate	SE	p	b-Estimate	SE	p
Gf-composite	**0.154**	**0.07**	**0.029**	0.139	0.081	0.086
WM-composite	0.098	0.07	0.16	**0.225**	**0.08**	**0.006**
Working memory	0.09	0.07	0.198	**0.228**	0.08	**0.005**
Verbal/spatial	**0.208**	**0.07**	**0.003**	0.057	0.081	0.483
Quantitative	0.001	0.069	0.984	**0.15**	**0.079**	**0.058**

Table 6
Gf factor description by sex.

	Males			Females		
	Working memory	Verbal/spatial	Quantitative	Working memory	Quantitative/spatial	Verbal
BOMAT	0.11	**0.74**	0.17	0.01	**0.67**	0.49
Number Series	0	0.03	**0.97**	0.04	**0.91**	−0.18
Letter Sets	−0.03	**0.93**	−0.04	0.07	−0.08	**0.9**
Reading span	**0.74**	0.17	−0.14	**0.64**	0.04	0.15
Rotation span	**0.88**	−0.14	0.09	**0.89**	−0.04	−0.02
Symmetry span	**0.78**	0.1	0.01	**0.82**	0.05	−0.02
Proportion variance explained	0.33	0.25	0.17	0.32	0.22	0.19

may require tests and analysis techniques (e.g., exploratory factor analysis) sensitive to identifying various sub-domains of intelligence.

Varieties of intelligence have been psychometrically described within several frameworks and taxonomies. The Cattel–Horn–Carrol (CHC) framework (Carroll, 1993; Horn and Noll, 1997; McGrew, 2009) is perhaps the most widely accepted psychometric-based theory of intelligence and posits at least nine broad domains of intelligence (including Gf), with dozens of narrow domains nested under them. Our factor analysis reveals that tests of quantitative and verbal reasoning split into separate factors, which is consistent with the specification of quantitative reasoning as a narrow ability feeding into Gf in the CHC framework (McGrew, 2009). Moreover, re-analysis of the WAIS-IV intelligence battery using a 5-factor (instead of 4-factor) structure (Benson et al., 2010; van Aken et al., 2015; Weiss et al., 2013) matches this psychometric description, revealing a quantitative reasoning sub-factor nested beneath Gf (Benson et al., 2010; Weiss et al., 2013).

The exploratory factor analysis we present here bears some resemblance to the CHC framework: BOMAT and Letter Sets together may resemble a more broad Gf factor combining verbal and spatial reasoning (this factor accounts for the most variance in our Gf test battery); Number Series primarily contributes to a quantitative reasoning factor (although not explicitly nested beneath the first factor, it accounts for less total variance in the Gf test battery). In the full sample, BOMAT loads primarily with Letter Sets, but also with Number Series (Table 3). This is consistent with previous results demonstrating that the Raven's Progressive Matrices (RPM), which is similar to the BOMAT, shares variance with both Letter Sets (Hambrick, 2003) and Number Series (Ackerman et al., 2002). Because the BOMAT has shared variance with both our spatial and quantitative reasoning components, it is likely that both are necessary for performing this test. Upon splitting our sample between males and females, BOMAT loads less on quantitative reasoning in males, but more on quantitative reasoning in females. Previous studies suggest that males perform better on quantitative reasoning (Geary et al., 2000), whereas females exhibit an advantage in verbal reasoning (Halpern, 2013). These relative strengths may even exist on an androgynous continuum, in which the degree of testosterone exposure in infancy predicts whether the pattern of cognitive performance is more male-like with higher quantitative performance, or more female-like with higher verbal performance (Luxen and Buunk, 2005). In our sample, although males showed an advantage in quantitative reasoning ability, results from both sexes demonstrate that the quantitative reasoning is positively correlated with brain volume, specifically gray matter (GM) volume. Some research has reported evidence that intelligence is more strongly related to gray matter structure in males and white matter structure in females (Gur et al., 1999; Haier et al., 2005); however, our data did not reveal this pattern of results, and instead suggest that the largest magnitude of correlation is between quantitative reasoning and GM volume for both males and females (Table 7).

While the CHC framework describes domains of intelligence derived from a broad array of tests, sub-domains of Gf have been previously characterized through factor analysis of item-level responses on the RPM. In addition to a perceptual or Gestalt factor, two distinct analytical factors have been identified (Lynn and Irwing, 2004; Mackintosh and Bennett, 2005). The first of these factors relates to verbal-analytic reasoning and exhibits no sex differences. The second analytic factor exhibits a male advantage (Mackintosh and Bennett, 2005). Interestingly, performance on this factor correlates with math ability in males but not in females, suggesting that males may solve these particular problems by employing specific cognitive processes related to mathematical reasoning or translation of problems into mathematical terminology, whereas females may employ more general cognitive abilities (Plaisted et al., 2011). These findings suggest that the verbal-analytic factor may provide

Table 7
Bivariate correlations between Gf and brain biomarkers by sex. All p-values are two-tailed.

		NAAt-corrected	Brain volume	GM volume	WM volume
Males n = 121	Gf-composite	0.144 (p = 0.115)	0.129 (p = 0.16)	0.143 (p = 0.117)	**0.103 (p = 0.26)**
	Working memory-composite	0.111 (p = 0.226)	**0.178 (p = 0.051)**	**0.178 (p = 0.051)**	**0.164 (p = 0.072)**
	Working memory	0.093 (p = 0.313)	**0.176 (p = 0.053)**	**0.178 (p = 0.051)**	**0.161 (p = 0.078)**
	Verbal/spatial	**0.212 (p = 0.019)**	0.045 (p = 0.625)	0.047 (p = 0.606)	0.039 (p = 0.673)
	Quantitative	−0.026 (p = 0.775)	**0.186 (p = 0.041)**	**0.212 (p = 0.02)**	0.144 (0.115)
Females n = 90	Gf-composite	0.12 (p = 0.258)	0.143 (p = 0.178)	0.141 (p = 0.186)	0.133 (p = 0.212)
	Working memory-composite	0.056 (p = 0.598)	**0.251 (p = 0.017)**	**0.239 (p = 0.023)**	**0.241 (p = 0.022)**
	Working memory	0.048 (p = 0.652)	**0.249 (p = 0.018)**	**0.239 (p = 0.023)**	**0.238 (p = 0.024)**
	Verbal	**0.195 (p = 0.065)**	0.068 (p = 0.527)	0.026 (p = 0.808)	0.108 (p = 0.311)
	Quantitative/spatial	0.018 (p = 0.863)	0.143 (p = 0.178)	**0.175 (p = 0.099)**	0.095 (p = 0.372)

a more representative, global measurement of Gf, characterizing reasoning abilities that do not differ by sex (Plaisted et al., 2011).

A comparison between this literature and our results reveals similar patterns. As with the second analytic component (Mackintosh and Bennett, 2005), sex differences are observed in our sample for the Number Series test, which principally contributes to our quantitative reasoning Gf factor. Moreover, factor analysis for males only showed that performance on the Number Series test is distinct from performance on the BOMAT, whereas for females the Number Series test loaded with the BOMAT (Table 6). The BOMAT is closely related to the RPM, which is believed to be an accurate measurement of a general cognitive process that underscores intelligence (Jensen, 1998). These results therefore support the possibility that females may have been more likely to use general fluid reasoning processes to complete the Number Series test, whereas males may have employed more specific cognitive processes related to mathematical reasoning, though further experiments designed to specifically test this hypothesis are necessary.

In contrast to the Number Series test, the Letter Sets test and the BOMAT test exhibit no sex differences and contribute to the verbal/spatial reasoning component. This component resembles the verbal-analytic reasoning factor derived from the RPM, which does not exhibit sex differences and may reflect a more general cognitive process (Plaisted et al., 2011). In our data, we observe positive correlations between NAA and verbal/spatial reasoning; thus, NAA concentration may be sensitive to general cognitive processes that underscore fluid intelligence. In our data, brain volume correlated positively with the quantitative reasoning factor; thus, brain size may be more sensitive to specific cognitive processes that differ between sexes and that do not contribute significantly to general intelligence (Burgaleta et al., 2012).

Our measurements of NAA were taken from medial parietal gray matter and the posterior cingulate cortex (PCC). The current results suggest that the oxidative metabolism in the PCC may also be sensitive to fluid cognitive processes. The Parieto-Frontal Integration Theory of intelligence proposes a four-stage information processing model of intelligence (Colom et al., 2010; Jung & Haier, 2007). In the first two stages, sensory information is initially processed, integrated across modalities, and abstracted. In the final two stages the frontal and parietal regions interact to form hypotheses and form and inhibit responses. The relationship of the PCC to regions involved in these final two stages, in concert with our results demonstrating PCC NAA to predict fluid reasoning ability, suggest that oxidative metabolism in the PCC may be reflective of high functioning in the frontal–parietal network responsible for hypothesis and response formation. Furthermore, our results provide preliminary evidence that this network of brain regions may be less related to quantitative reasoning and more exclusively related to verbal and/or spatial reasoning processes (perhaps depending on sex). Recent work suggests that subcomponents of intelligence are indeed fractioned across different, unique brain networks (Hampshire et al., 2012).

Limitations and future directions

The additional domains of cognition measured by our battery may provide clues about the relationship of NAA to Gf. However, our battery did not include other cognitive faculties that may be of particular relevance for understanding the NAA–Gf relationship. For example, some evidence suggests that NAA may be sensitive to processing speed as measured by performance on timed tests (Jung et al., 1999), owing to its role as a precursor for myelin in WM (Moffett et al., 2007), though a recent report did not find support for this hypothesis (Patel et al., 2014). It is possible that different processing speed demands, or interactions between NAA concentration and processing speed demands may factor into our reported dissociation. Without processing speed measures, our data cannot account for this possibility; thus, future studies will benefit from including a wider array of tests to help account for these and other possibilities.

Future studies may also benefit from experiments designed to more fully characterize the water signal that is used to normalize the NAA signal. Such experiments are motivated by the fact that creatine and choline, which also show weak correlations with verbal/spatial reasoning, are normalized by the same water signal. To learn whether this water signal may be contributing to the observed correlations, future experiments can include direct measurements of the relaxation properties of the water and the volume of NMR-visible water in the voxel (Gasparovic et al., 2009), thereby avoiding the need to assume the same literature values for these properties in all subjects. Additionally, future studies will benefit from multi-voxel MRS techniques providing measurement of NAA in a variety of brain regions (Jung et al., 1999; Nikolaidis et al., 2016). This information should be combined with other types of imaging data including regionally specific MRI measures of volumetrics in brain regions specifically linked to intelligence, structural white matter integrity between such regions (via DTI), or perhaps functional connectivity between regions (via resting state fMRI) to disentangle how regional neurometabolites, structure, and function together support Gf.

Finally, studies seeking to better understand intelligence and its relationship to brain metabolites should investigate whether controlled interventions can produce changes in NAA concentration, and critically, whether these changes are associated with improvements in intelligence. Whether or not such interventions are efficacious may further inform the clinical relevancy of these biomarkers to diseases such as Alzheimer's disease and autism spectrum disorder, in which both brain volume and NAA concentrations are altered (Corrigan et al., 2013; Fox and Schott, 2004; Pfefferbaum et al., 1999; Piven et al., 1995).

Conclusion

To our knowledge, this is the largest study investigating the relationship between the concentration of the brain metabolite NAA and Gf. Despite years of research consistently demonstrating positive correlations between brain volume and intelligence, our results showed that NAA concentration and brain volume are dissociable biomarkers of facets of Gf independent of working memory: NAA positively correlated with verbal/spatial Gf whereas brain volume correlated with quantitative Gf. Additionally, this pattern of results was observed for males and females separately. This finding suggests that different brain biomarkers may uniquely reflect different aspects of human cognition, and that a complete understanding of the neurobiological bases of intelligence requires a complete characterization of such biomarkers.

Acknowledgments

The research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via Contract 2014-13121700004 to the University of Illinois at Urbana-Champaign (PI: Barbey). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

We thank Nancy Dodge, Holly Tracy, Tracy Henigman, and Courtney Allen for assistance in performing the experiments; we also thank Michael Newman for assistance with data processing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.05.037.

References

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. 2 (4), 433–459.

Ackerman, P.L., Beier, M.E., Boyle, M.O., 2002. Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. J. Exp. Psychol. Gen. 131 (4), 567.

Ackerman, P.L., Beier, M.E., Boyle, M.O., 2005. Working memory and intelligence: the same or different constructs? Psychol. Bull. 131 (1), 30.

Andrews-Hanna, J.R., Snyder, A.Z., Vincent, J.L., Lustig, C., Head, D., Raichle, M.E., Buckner, R.L., 2007. Disruption of large-scale brain systems in advanced aging. Neuron 56 (5), 924–935.

Aydin, K., Uysal, S., Yakut, A., Emiroglu, B., Yilmaz, F., 2012. N-Acetylaspartate concentration in corpus callosum is positively correlated with intelligence in adolescents. NeuroImage 59 (2), 1058–1064. http://dx.doi.org/10.1016/j.neuroimage.2011.08.114.

Baniqued, P.L., Kranz, M.B., Voss, M.W., Lee, H., Cosman, J.D., Severson, J., Kramer, A.F., 2014. Cognitive training with casual video games: points to consider. Front. Psychol. 4.

Barbey, A.K., Colom, R., Grafman, J., 2013a. Architecture of cognitive flexibility revealed by lesion mapping. NeuroImage 82, 547–554.

Barbey, A.K., Colom, R., Grafman, J., 2013b. Dorsolateral prefrontal contributions to human intelligence. Neuropsychologia 51 (7), 1361–1369.

Barbey, A.K., Colom, R., Paul, E.J., Grafman, J., 2014. Architecture of fluid intelligence and working memory revealed by lesion mapping. Brain Struct. Funct. 219 (2), 485–494.

Barker, P.B., Hearshen, D.O., Boska, M.D., 2001. Single-voxel proton MRS of the human brain at 1.5 T and 3.0 T. Magn. Reson. Med. 45 (5), 765–769.

Benson, N., Hulac, D.M., Kranzler, J.H., 2010. Independent examination of the Wechsler Adult Intelligence Scale — Fourth Edition (WAIS-IV): what does the WAIS-IV measure? Psychol. Assess. 22 (1), 121.

Brooks, W.M., Friedman, S.D., Stidley, C.A., 1999. Reproducibility of 1H-MRS in vivo. Magn. Reson. Med. 41 (1), 193–197.

Burgaleta, M., Head, K., Álvarez-Linera, J., Martínez, K., Escorial, S., Haier, R., Colom, R., 2012. Sex differences in brain volume are related to specific skills, not to general intelligence. Intelligence 40 (1), 60–68.

Carroll, J.B., 1993. Human Cognitive Abilities: A Survey of Factor-Analytic Studies. Cambridge University Press.

Cattell, R.B., 1971. Abilities: Their Structure, Growth, and Action. Houghton Mifflin, Boston.

Charlton, R.A., McIntyre, D.J.O., Howe, F.A., Morris, R.G., Markus, H.S., 2007. The relationship between white matter brain metabolites and cognition in normal aging: the GENIE study. Brain Res. 1164, 108–116.

Chuderski, A., 2013. When are fluid intelligence and working memory isomorphic and when are they not? Intelligence 41 (4), 244–262.

Colom, R., Flores-Mendoza, C.E., 2007. Intelligence predicts scholastic achievement irrespective of SES factors: evidence from Brazil. Intelligence 35 (3), 243–251.

Colom, R., Karama, S., Jung, R.E., Haier, R.J., 2010. Human intelligence and brain networks. Dialogues Clin. Neurosci. 12 (4), 489–501.

Colom, R., Privado, J., García, L.F., Estrada, E., Cuevas, L., Shih, P.C., 2015. Fluid intelligence and working memory capacity: is the time for working on intelligence problems relevant for explaining their large relationship? Personal. Individ. Differ. 79, 75–80.

Conway, A.R., Kane, M.J., Engle, R.W., 2003. Working memory capacity and its relation to general intelligence. Trends Cogn. Sci. 7 (12), 547–552.

Corrigan, N.M., Shaw, D.W., Estes, A.M., Richards, T.L., Munson, J., Friedman, S.D., ... Dager, S.R., 2013. Atypical developmental patterns of brain chemistry in children with autism spectrum disorder. JAMA Psychiatry 70 (9), 964–974.

Edden, R.A., Pomper, M.G., Barker, P.B., 2007. In vivo differentiation of N-acetyl aspartyl glutamate from N-acetyl aspartate at 3 Tesla. Magn. Reson. Med. 57 (6), 977–982.

Ekstrom, R.B., French, J.W., Harman, H., Derman, D., 1976. Kit of Factor-Referenced Cognitive Tests. rev. ed. Educational Testing Service, Princeton, NJ.

Ferguson, K.J., MacLullich, A.M.J., Marshall, I., Deary, I.J., Starr, J.M., Seckl, J.R., Wardlaw, J.M., 2002. Magnetic resonance spectroscopy and cognitive function in healthy elderly men. Brain 125 (12), 2743–2749.

Foster, J.L., Shipstead, Z., Harrison, T.L., Hicks, K.L., Redick, T.S., Engle, R.W., 2015. Shortened complex span tasks can reliably measure working memory capacity. Mem. Cogn. 43 (2), 226–236.

Fox, N.C., Schott, J.M., 2004. Imaging cerebral atrophy: normal ageing to Alzheimer's disease. Lancet 363 (9406), 392–394.

Gasparovic, C., Neeb, H., Feis, D.L., Damaraju, E., Chen, H., Doty, M.J., ... Shah, N.J., 2009. Quantitative spectroscopic imaging with in situ measurements of tissue water T1, T2, and density. Magn. Reson. Med. 62 (3), 583–590.

Gasparovic, C., Song, T., Devier, D., Bockholt, H.J., Caprihan, A., Mullins, P.G., ... Morrison, L.A., 2006. Use of tissue water as a concentration reference for proton spectroscopic imaging. Magn. Reson. Med. 55 (6), 1219–1226.

Geary, D.C., Saults, S.J., Liu, F., Hoard, M.K., 2000. Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. J. Exp. Child Psychol. 77 (4), 337–353.

Gignac, G., Vernon, P.A., Wickett, J.C., 2003. Factors influencing the relationship between brain size and intelligence. In: Nyborg, H. (Ed.), The Scientific Study of General Intelligence: Tribute to Arthur R. Jensen. Pergamon Press, pp. 93–106.

Giménez, M., Junqué, C., Narberhaus, A., Caldú, X., Segarra, D., Vendrell, P., Bargalló, N., Mercader, J.M., 2004. Medial temporal MR spectroscopy is related to memory performance in normal adolescent subjects. Neuroreport 15 (4), 703–707.

Gottfredson, L., Saklofske, D.H., 2009. Intelligence: foundations and issues in assessment. Can. Psychol./Psychol. Can. 50 (3), 183.

Gray, J.R., Chabris, C.F., Braver, T.S., 2003. Neural mechanisms of general fluid intelligence. Nat. Neurosci. 6 (3), 316–322.

Gray, J.R., Thompson, P.M., 2004. Neurobiology of intelligence: science and ethics. Nat. Rev. Neurosci. 5 (6), 471–482.

Gur, R.C., Turetsky, B.I., Matsui, M., Yan, M., Bilker, W., Hughett, P., Gur, R.E., 1999. Sex differences in brain gray and white matter in healthy young adults: correlations with cognitive performance. J. Neurosci. 19 (10), 4065–4072.

Gussew, A., Erdtel, M., Hiepe, P., Rzanny, R., Reichenbach, J.R., 2012. Absolute quantitation of brain metabolites with respect to heterogeneous tissue compositions in 1H-MR spectroscopic volumes. MAGMA 25 (5), 321–333.

Haier, R.J., Chueh, D., Touchette, P., Lott, I., Buchsbaum, M.S., MacMillan, D., ... Sosa, E., 1995. Brain size and cerebral glucose metabolic rate in nonspecific mental retardation and Down syndrome. Intelligence 20 (2), 191–210.

Haier, R.J., Jung, R.E., Yeo, R.A., Head, K., Alkire, M.T., 2005. The neuroanatomy of general intelligence: sex matters. NeuroImage 25 (1), 320–327.

Halpern, D.F., 2013. Sex Differences in Cognitive Abilities. Psychology Press.

Hambrick, D.Z., 2003. Why are some people more knowledgeable than others? A longitudinal study of knowledge acquisition. Mem. Cogn. 31 (6), 902–917.

Hampshire, A., Highfield, R.R., Parkin, B.L., Owen, A.M., 2012. Fractionating human intelligence. Neuron 76 (6), 1225–1237.

Harrison, T.L., Shipstead, Z., Hicks, K.L., Hambrick, D.Z., Redick, T.S., Engle, R.W., 2013. Working memory training may increase working memory capacity but not fluid intelligence. Psychol. Sci. (0956797613492984).

Haznedar, M.M., Buchsbaum, M.S., Hazlett, E.A., Shihabuddin, L., New, A., Siever, L.J., 2004. Cingulate gyrus volume and metabolism in the schizophrenia spectrum. Schizophr. Res. 71, 249–262.

Horn, J.L., Cattell, R.B., 1966. Refinement and test of the theory of fluid and crystallized general intelligences. J. Educ. Psychol. 57 (5), 253.

Horn, J.L., Cattell, R.B., 1967. Age differences in fluid and crystallized intelligence. Acta Psychol. 26, 107–129.

Horn, J.L., Noll, J., 1997. Human cognitive capabilities: Gf–Gc theory. In: Flanagan, D.P., Genshaft, J.L., Harrison, P.L. (Eds.), Contemporary Intellectual Assessment: Theories, Tests and Issues. Guilford, New York, pp. 53–91.

Hossiep, R., Turck, D., Hasella, M., 1999. Bochumer Matrizentest (BOMAT) Advanced. Hogrefe, Göttingen.

Irwing, P., Lynn, R., 2005. Sex differences in means and variability on the progressive matrices in university students: a meta-analysis. Br. J. Psychol. 96, 505–524.

Ivanovic, D.M., Leiva, B.P., Pérez, H.T., Olivares, M.G., Diaz, N.S., Urrutia, M.S.C., et al., 2004. Head size and intelligence, learning, nutritional status and brain development: head, IQ, learning, nutrition and brain. Neuropsychologia 42 (8), 1118–1131.

Jaeggi, S.M., Buschkuehl, M., Jonides, J., Perrig, W.J., 2008. Improving fluid intelligence with training on working memory. Proc. Natl. Acad. Sci. U. S. A. 105 (19), 6829–6833.

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. Fsl. NeuroImage 62 (2), 782–790.

Jensen, A.R., 1980. Bias in Mental Testing. The Free Press, New York.

Jensen, A.R., 1998. The g Factor: The Science of Mental Ability. Praeger Westport, CT.

Jolliffe, I., 2002. Principal Component Analysis. Wiley Online Library.

Jung, R.E., Haier, R.J., 2007. The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. Behav. Brain Sci. 30 (02), 135–154.

Jung, R.E., Brooks, W.M., Yeo, R.A., Chiulli, S.J., Weers, D.C., Sibbitt, W.L., 1999. Biochemical markers of intelligence: a proton MR spectroscopy study of normal human brain. Proc. R. Soc. Lond. B Biol. Sci. 266 (1426), 1375–1379.

Jung, R.E., Gasparovic, C., Chavez, R.S., Caprihan, A., Barrow, R., Yeo, R.A., 2009. Imaging intelligence with proton magnetic resonance spectroscopy. Intelligence 37 (2), 192–198.

Jung, R.E., Haier, R.J., Yeo, R.A., Rowland, L.M., Petropoulos, H., Levine, A.S., ... Brooks, W.M., 2005. Sex differences in N-acetylaspartate correlates of general intelligence: an 1 H-MRS study of normal human brain. NeuroImage 26 (3), 965–972.

Kane, M.J., Engle, R.W., 2002. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. Psychon. Bull. Rev. 9 (4), 637–671.

Kane, M.J., Hambrick, D.Z., Conway, A.R., 2005. Working Memory Capacity and Fluid Intelligence are Strongly Related Constructs: Comment on Ackerman, Beier, and Boyle (2005).

Kochunov, P., Coyle, T., Lancaster, J., Robin, D.A., Hardies, J., Kochunov, V., ... Null, M., 2010. Processing speed is correlated with cerebral health markers in the frontal lobes as quantified by neuroimaging. NeuroImage 49 (2), 1190–1199.

Kyllonen, P.C., Christal, R.E., 1990. Reasoning ability is (little more than) working-memory capacity? Intelligence 14 (4), 389–433.

Leech, R., Sharp, D.J., 2014. The role of the posterior cingulate cortex in cognition and disease. Brain J. Neurol. 137, 12–32. http://dx.doi.org/10.1093/brain/awt162.

Luders, E., Narr, K.L., Thompson, P.M., Toga, A.W., 2009. Neuroanatomical correlates of intelligence. Intelligence 37 (2), 156–163.

Luxen, M.F., Buunk, B.P., 2005. Second-to-fourth digit ratio related to verbal and numerical intelligence and the Big Five. Personal. Individ. Differ. 39 (5), 959–966.

Lynn, R., Irwing, P., 2004. Sex differences on the progressive matrices: a meta-analysis. Intelligence 32 (5), 481–498.

Mackintosh, N.J., Bennett, E.S., 2005. What do Raven's matrices measure? An analysis in terms of sex differences. Intelligence 33 (6), 663–674.

Martínez, K., Burgaleta, M., Román, F.J., Escorial, S., Shih, P.C., Quiroga, M.Á, Colom, R., 2011. Can fluid intelligence be reduced to 'simple'short-term storage? Intelligence 39 (6), 473–480.

McDaniel, M.A., 2005. Big-brained people are smarter: a meta-analysis of the relationship between in vivo brain volume and intelligence. Intelligence 33 (4), 337–346.

McGrew, K.S., 2009. CHC theory and the human cognitive abilities project: standing on the shoulders of the giants of psychometric intelligence research. Intelligence 37 (1), 1–10.

Minoshima, S., Giordani, B., Berent, S., Frey, K.A., Foster, N.L., Kuhl, D.E., 1997. Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. Ann. Neurol. 42, 85–94.

Moffett, J.R., Ross, B., Arun, P., Madhavarao, C.N., Namboodiri, A.M.A., 2007. N-Acetylaspartate in the CNS: from neurodiagnostics to neurobiology. Prog. Neurobiol. 81 (2), 89–131.
Moody, D.E., 2009. Can intelligence be increased by training on a task of working memory? Intelligence 37 (4), 327–328.
Nakashima, T., Nakayama, N., Miwa, K., Okumura, A., Soeda, A., Iwama, T., 2007. Focal brain glucose hypometabolism in patients with neuropsychologic deficits after diffuse axonal injury. AJNR 28, 236–242.
Neisser, U., Boodoo, G., Bouchard Jr., T.J., Boykin, A.W., Brody, N., Ceci, S.J., et al., 1996. Intelligence: knowns and unknowns. Am. Psychol. 51 (2), 77.
Nikolaidis, A., Baniqued, P.L., Kranz, M.B., Scavuzzo, C.J., Barbey, A.K., Kramer, A.F., Larsen, R.J., 2016. Multivariate associations of fluid intelligence and NAA. Cereb. Cortex, bhw070.
Oswald, F.L., McAbee, S.T., Redick, T.S., Hambrick, D.Z., 2015. The development of a short domain-general measure of working memory capacity. Behav. Res. Methods 47 (4), 1343–1355.
Patel, T., Blyth, J.C., Griffiths, G., Kelly, D., Talcott, J.B., 2014. Moderate relationships between NAA and cognitive ability in healthy adults: implications for cognitive spectroscopy. Front. Hum. Neurosci. 8.
Pfefferbaum, A., Adalsteinsson, E., Spielman, D., Sullivan, E.V., Lim, K.O., 1999. In vivo spectroscopic quantification of the N-acetyl moiety, creatine, and choline from large volumes of brain gray and white matter: effects of normal aging. Magn. Reson. Med. 41 (2), 276–284.
Pfleiderer, B., Ohrmann, P., Suslow, T., Wolgast, M., Gerlach, A.L., Heindel, W., Michael, N., 2004. N-acetylaspartate levels of left frontal cortex are associated with verbal intelligence in women but not in men: a proton magnetic resonance spectroscopy study. Neuroscience 123 (4), 1053–1058.
Piven, J., Arndt, S., Bailey, J., Havercamp, S., Andreasen, N.C., Palmer, P., 1995. An MRI study of brain size in autism. Am. J. Psychiatr. 152 (8), 1145–1149.
Plaisted, K., Bell, S., Mackintosh, N.J., 2011. The role of mathematical skill in sex differences on Raven's matrices. Personal. Individ. Differ. 51 (5), 562–565.
Provencher, S.W., 1993. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. Magn. Reson. Med. 30 (6), 672–679.
Ritchie, S.J., Booth, T., Hernández, M.D.C.V., Corley, J., Maniega, S.M., Gow, A.J., ... Bastin, M.E., 2015. Beyond a bigger brain: multivariable structural brain imaging and intelligence. Intelligence 51, 47–56.
Ross, A.J., Sachdev, P.S., 2004. Magnetic resonance spectroscopy in cognitive research. Brain Res. Rev. 44 (2), 83–102.
Ross, A.J., Sachdev, P.S., Wen, W., Valenzuela, M.J., Brodaty, H., 2005. Cognitive correlates of 1 H MRS measures in the healthy elderly brain. Brain Res. Bull. 66 (1), 9–16.
Rushton, J.P., Ankney, C.D., 2009. Whole brain size and general mental ability: a review. Int. J. Neurosci. 119 (5), 692–732.
Schmithorst, V.J., 2009. Developmental sex differences in the relation of neuroanatomical connectivity to intelligence. Intelligence 37 (2), 164–173.
Schönbrodt, F.D., Perugini, M., 2013. At what sample size do correlations stabilize? J. Res. Pers. 47 (5), 609–612.
Spearman, C., 1904. "General intelligence," objectively determined and measured. Am. J. Psychol. 15 (2), 201–292.
Spearman, C., 1927. The Abilities of Man. Macmillan, New York.
Tabachnick, B.G., Fidell, L.S., 2007. Using Multivariate Statistics. fifth ed Pearson Allyn & Bacon, Upper Saddle River, NJ.
Terpstra, M., Cheong, I., Lyu, T., Deelchand, D.K., Emir, U.E., Bednařík, P., ... Öz, G., 2015. Test-retest reproducibility of neurochemical profiles with short-echo, single-voxel MR spectroscopy at 3T and 7T. Magn. Reson. Med.
Thurstone, L.L., 1938. Primary Mental Abilities. University of Chicago Press, Chicago.
van Aken, L., van der Heijden, P.T., van der Veld, W.M., Hermans, L., Kessels, R.P., Egger, J.I., 2015. Representation of the Cattell–Horn–Carroll theory of cognitive abilities in the factor structure of the Dutch-Language version of the WAIS-IV. Assessment (1073191115607973).
Weiss, L.G., Keith, T.Z., Zhu, J., Chen, H., 2013. WAIS-IV and clinical validation of the four-and five-factor interpretative approaches. J. Psychoeduc. Assess. (0734282913478030).
Wickett, J.C., Vernon, P.A., Lee, D.H., 2000. Relationships between factors of intelligence and brain volume. Personal. Individ. Differ. 29 (6), 1095–1122.
Witelson, S.F., Beresh, H., Kigar, D.L., 2006. Intelligence and brain size in 100 postmortem brains: sex, lateralization and age factors. Brain 129 (2), 386–398.
Yeo, R.A., Hill, D., Campbell, R., Vigil, J., Brooks, W.M., 2000. Developmental instability and working memory ability in children: a magnetic resonance spectroscopy investigation. Dev. Neuropsychol. 17 (2), 143–159.
Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation–maximization algorithm. IEEE Trans. Med. Imaging 20 (1), 45–57.