Received: 20 January 2016 | Accepted: 1 December 2016
DOI: 10.1111/desc.12545

SHORT REPORT

WILEY Developmental Science

Neuroanatomical correlates of performance in a state-wide test of math achievement

Eric D. Wilkey[1,2] | Laurie E. Cutting[1,2] | Gavin R. Price[1,2]

[1]Peabody College of Education and Human Development, Vanderbilt University, Nashville, TN, USA

[2]Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN, USA

Correspondence
Gavin R. Price, Department of Psychology & Human Development, Peabody College, Vanderbilt University, 230 Appleton Place, Nashville, TN 37203, USA.
Email: gavin.price@vanderbilt.edu

Funding information
National Institutes of Health, Grant/Award Number: P30 HD015052; National Center for Research Resources, Grant/Award Number: CTSA RR024975; National Institute of Child Health and Human Development, Grant/Award Number: RO1 HD044073 and RO1 HD067254

Abstract

The development of math skills is a critical component of early education and a strong indicator of later school and economic success. Recent research utilizing population-normed, standardized measures of math achievement suggest that structural and functional integrity of parietal regions, especially the intraparietal sulcus, are closely related to the development of math skills. However, it is unknown how these findings relate to in-school math learning. The present study is the first to address this issue by investigating the relationship between regional differences in grey matter (GM) volume and performance in grade-level mathematics as measured by a state-wide, school-based test of math achievement (TCAP math) in children from 3rd to 8th grade. Results show that increased GM volume in the bilateral hippocampal formation and the right inferior frontal gyrus, regions associated with learning and memory, is associated with higher TCAP math scores. Secondary analyses revealed that GM volume in the left angular gyrus had a stronger relationship to TCAP math in grades 3–4 than in grades 5–8 while the relationship between GM volume in the left inferior frontal gyrus and TCAP math was stronger for grades 5–8. These results suggest that the neuroanatomical architecture related to in-school math achievement differs from that related to math achievement measured by standardized tests, and that the most related neural structures differ as a function of grade level. We suggest, therefore, that the use of school-relevant outcome measures is critical if neuroscience is to bridge the gap to education.

RESEARCH HIGHLIGHTS

- This study investigates relationship between whole-brain grey matter (GM) volume and performance on the math section of a state-wide, school-based test of math achievement in grades 3–8.
- Greater GM volume in the bilateral hippocampus and right inferior frontal gyrus was associated with higher performance on grade-level tests of math achievement.
- For grades 3–4, the relationship between GM volume in the left angular gyrus and math was stronger; for grades 5–8, the left inferior frontal gyrus.
- GM in regions of the brain associated with generalized learning and memory are more closely related to in-school math achievement than areas of the brain associated with magnitude processing (i.e., intraparietal sulcus).

1 | INTRODUCTION

Early math skills are a strong predictor of an individual's academic achievement (Duncan et al., 2007), college entry (Bynner & Parsons, 1997), employment status (Ritchie & Bates, 2013), and physical and mental health (Bynner & Parsons, 2006). On a societal level, small increases in a nation's numeracy rate are related to observable increases in GDP (OECD, 2010). With this in mind, many governments have dedicated resources and implemented new initiatives to enhance math education (British Council, 2010; White House, Office of the Press Secretary, 2010). One broad trend in these efforts is the application of neuroscience to questions of education. There are now international conferences, professional societies, graduate education programs, and scientific journals dedicated to this burgeoning field of 'educational neuroscience'. This field is defined by the belief that

TABLE 1 Studies of neuroanatomical differences underlying math competence

Study	Age (*SD*)	Study design	Anatomical metric	Index of math competency	Type of math math measure	Main findings
Isaacs et al. 2001	15.83 (*1.1*)	LMA vs. HMA	GMV	WOND	Norm-based	LMA show lower GMV in left IPS.
Han et al. 2013	10.8 (*0.4*)	LMA vs. HMA	Morphometric differences	WRAT (arithmetic)	Norm-based	Anatomical differences in the L occipital-temporal cortex, L orbital-frontal cortex, and R insular cortex relates to LMA vs. HMA.
Rotzer et al. 2008	9.3 (*0.2*)	LMA vs. HMA	GMV	ZAREKI-R (Num. Proc. & Calculation)	Norm-based	LMA showed reduced GMV in R IPS, anterior cingulum, L IFG, and bilateral middle frontal gyri.
Rykhlevskaia et al. 2009	8.8 (*0.7*)	LMA vs. HMA	GMV	WIAT (Num. Ops. & Math Reasoning)	Norm-based	LMA show reduced GMV in bilateral superior parietal lobe, IPS, fusiform gyrus, para-hippocampal gyrus, and R anterior temporal cortex.
Ranpura et al. 2013	8–14	LMA vs. HMA	CSA, CT, and GMV	WOND	Norm-based	LMA show reduced CSA in subcentral gyri, reduced CT in L temporal and R IFG, and less GMV in R parrahippocampal gyrus and R parietal lobe.
Cappelletti & Price 2014	22–74	LMA vs. HMA	GMV	(1) Dyscalculia Screener (2) Graded Difficulty Arithmetic	(1) Norm-based (2) Norm-based	LMA show reduced GMV in R IPS.
Lubin et al. 2013	10 (*0.6*)	High vs. Low Number Skills	GMV	(1) Number Line Estimation (2) Visual Numeracy Estimation	(1) Experimental (2) Experimental	LMA show reduced GMV in L IPS, bilateral Angular Gyrus. And occipito-temporal areas.
Starke et al. 2013	6.3–7.9	TA	GMV	(1) TEDI-MATH (2) Number Comparison Task	(1) Norm-based (2) Experimental	Positive relationship between GMV and numerical distance effect in multiple parietal regions and superior temporal gyrus.
Li et al. 2013	10.48 (*0.4*)	TA	GMV & FA	WISC (arithmetic)	Norm-based	Arithmetic scores positively correlated with GMV in R IPS.
Price et al. 2016	7.44 (*0.3*)	TA	GMV	WCJ-III (App. Problems & Calculation)	Norm-based	Composite math scores positively correlated with GMV in L IPS.
Supekar et al. 2013	8.7 (*0.1*)	TA	GMV	(1) Arithmetic Verification (2) Arithmetic Production	Linked to instruction outcomes	GMV in R hippocampus predicted math performance gains.

Note. LMA = Low Math Ability; HMA = High Math Ability; TA = Typically Achieving, individual differences; GMV = Grey Matter Volume; CSA = Cortical Surface Area; CT = Cortical Thickness; WOND = Wechsler Objective Numeric Dimensions Test; WRAT = Wide Range Achievement Test; ZAREKI-R = Neuropsychological Test Battery for Number Processing and Calculation in Children, Revised; WIAT = Wechsler Individual Achievement Test; WCJ-III = Woodcock Johnson III understanding the neurocognitive foundations of academic skills will contribute to improvements in the efficacy of educational interventions and pedagogical methods, and in turn, lead to better utilization of resources dedicated to education.

As a result of applying neuroscience methods to educationally relevant topics, several advances have been made in our understanding of the neurocognitive foundations of math-related abilities. Functional magnetic resonance imaging (fMRI) has provided detailed information about the neural correlates of basic numeracy skills (Bulthé, De Smedt, & Op de Beeck, 2014; Hubbard, Piazza, Pinel, & Dehaene, 2005; Price & Ansari, 2011), specific deficits in math learning disabilities (Price, Holloway, Räsänen, Vesterinen, & Ansari, 2007; Rosenberg-Lee et al., 2015), and individual variability in response to basic numeracy training (Kucian et al., 2011). Specifically, a significant body of research indicates that populations of neurons in the intraparietal sulcus (IPS) support the processing of numerical magnitude (Ansari & Dhital, 2006; Cantlon & Brannon, 2006; Cantlon, Brannon, Carter, & Pelphrey, 2006). Further, individual variability in the efficiency of this processing, measured behaviorally and with patterns of neural activation, is related to individual differences in math performance (Bugden & Ansari, 2011; Halberda, Mazzocco, & Feigenson, 2008; Mazzocco, Feigenson, & Halberda, 2011). In addition, activation of the left angular gyrus (AG) has been associated with the processing of numerical symbols (Price & Ansari, 2011), fluent retrieval of arithmetic facts (Delazer et al., 2005), and individual differences in math competence (Grabner et al., 2007). Several anatomical studies have identified structural abnormalities such as reduced grey matter (GM), abnormal sulcul geometry, and reduced white matter integrity in the bilateral IPS and surrounding white matter tracts in children with mathematical learning difficulties (Han et al., 2008; Isaacs, Edmonds, Lucas, & Gadian, 2001; Lubin et al., 2013; Rykhlevskaia, Uddin, Kondos, & Menon, 2009; Starke et al., 2013). More recently, increased GM volume of the left IPS has been shown to correlate with higher math scores in typically developing populations (Li, Hu, Wang, Weng, & Chen, 2013; Price, Wilkey, Yeo, & Cutting, 2016). Taken together, the extant literature suggests that the functional and structural integrity of domain-specific numerical processing regions in the superior and inferior parietal lobes play a key role in the development of math competence.

However, questions still remain regarding the extent to which neuroscience research methods can 'bridge the gap' to educational relevance (Bruer, 1997, 2006). One source of concern in this respect is the relationship between measures of math performance employed in previous studies and content being learned in the classroom in differing educational contexts. The most frequently used measures of math performance, such as the Woodcock Johnson Test of Achievement (Bugden, Price, McLean, & Ansari, 2012) or the Test of Early Mathematics Ability (Emerson & Cantlon, 2012), index individual achievement levels on a wide range of math skills and may include only a small portion of content that was covered in a student's classroom during a given year. These standardized measures of achievement provide important metrics of how a student's math competency compares to their peers in absolute terms (i.e., what is an individual's maximum level of math skill on a wide range of math-related content?), but they are designed to be administered to a wide age/ability range. By necessity, they are not tailored to be sensitive to a student's educational setting. In contrast, one recent study investigated the functional and anatomical neural predictors of response to an 8-week arithmetic intervention (Supekar et al., 2013). They found that volume and intrinsic functional connectivity of the right hippocampus predicted gains in arithmetic competence better than volume or connectivity in the IPS or AG. In other words, when they tailored their outcome measure to be sensitive to fluency and strategic efficiency of the math skill being taught, a region typically associated with memory encoding and retrieval (Eichenbaum, 2000) predicted gains better than regions typically associated with number processing or arithmetic. In comparison to previous studies of the neuroanatomical correlates of math competence (Table 1), the results of Supekar et al. (2013) suggest that the neural correlates of math competence may vary depending on the proximity of the outcome measure to the educational experience. However, the outcome measure used in that study was experimenter defined, albeit closely tied to the learning experience, leaving open the question of whether school-based math measures engage the same neurocognitive mechanisms as standardized math measures. If the effort to apply neuroscience research findings to education is to achieve the goal of having a direct impact on curriculum and pedagogy, measures of math performance used in research must be closely tied to what is happening in school. If researchers only use measures of math achievement that are distally related to classroom environments, it will remain difficult for neuroscience to bridge the gap to educational application. The present study addresses this issue by assessing the relationship between regional GM volume and scores on the Tennessee Comprehensive Achievement Program (TCAP) math subtest in a cross-sectional sample of students in grades 3–8. To the best of our knowledge this is the first study to relate brain structure to math performance utilizing a school-based test of math achievement.

The TCAP is a measure of grade-level concept mastery closely related to classroom learning in the preceding school year. It is used as a metric of student achievement and is incorporated into the student's year-end grade. As such, information gathered from the TCAP can influence students' courses and educational opportunities in the following years. Based on the previous functional and structural neuroimaging studies of the correlates of math achievement, two hypotheses are investigated. If performance on school-based measures of math competence is driven by domain-general cognitive abilities, such as executive functioning or memory formation, TCAP scores would be expected to correlate with volumetric differences in the prefrontal cortex, and the hippocampal formation. If, however, variability in structures associated with the processing of numerical magnitude and verbally represented numerical information are driving school math achievement, TCAP scores should correlate with differences in the bilateral IPS or AG. It is important to note that these hypotheses are not mutually exclusive, and a pattern of overlap and dissociation is possible, reflecting shared and distinct neurocognitive mechanisms underlying school-based measures of math competence.

TABLE 2 Descriptive statistics of student sample

	Mean	*SD*	Range
Age at scan (years)	11.47	1.14	10.06–14.77
Grade	4.7	1.14	3–8
Average months from TCAP to scan	8.2	3.6	1.5–13.5
TCAP math (*z*-score)	0.52	0.86	−1.22–2.06
TCAP reading (*z*-score)	0.56	0.96	−2.03–2.51
Performance IQ (WASI)*	105.5	13.4	77–134
Verbal IQ (WASI)	111.9	14.6	70–138
Full Scale IQ (WASI)*	109.9	13.6	78–136

Note. *SD*, standard deviation; TCAP, Tennessee Comprehensive Assessment Program; WASI, Wechsler Abbreviated Scale of Intelligence.
*$n = 48$ for Performance IQ and Full Scale IQ.

TABLE 3 Pearson *r*-values for bivariate correlations between academic and cognitive measures

Measure ($n = 49$)	1	2	3	4
1. TCAP math				
2. TCAP reading	.758**			
3. Performance IQ*	.639**	.389**		
4. Verbal IQ	.609**	.642**	.454**	
5. Full Scale IQ*	.720**	.603**	.840**	.862**

*$n = 48$ for Performance IQ and Full Scale IQ.
**indicates $p < .001$.

2 | METHODS

2.1 | Participants

The present data were collected as part of a larger scale, cross-sectional study investigating the development of reading comprehension (Aboud, Bailey, Petrill, & Cutting, 2016). The current sample ($n = 49$) comprises participants for whom end-of-year state tests of math achievement were available for the grade completed prior to MRI scans. The following exclusion criteria were applied prior to recruitment: (1) previous diagnosis of intellectual disability; (2) known, uncorrectable visual impairment; (3) treatment of any psychiatric disorder (other than ADHD) with psychotropic medications; (4) history of known neurological disorder (e.g., epilepsy, spina bifida, cerebral palsy, traumatic brain injury); (5) documented hearing impairment greater than or equal to 25 dB loss in either ear; (6) medical contraindication to MRI procedures (e.g., metal devices); (7) the history of or presence of a pervasive developmental disorder; and (8) if during testing, parental responses from the Diagnostic Interview for Children and Adolescents-IV (DICA-IV; Reich, Leacock, & Shanfeld, 1997) indicate the presence of any severe psychiatric diagnoses, including major depression, bipolar disorders, and conduct disorder. Individuals meeting criteria for ADHD, oppositional defiant disorder (ODD), adjustment disorder, and mild depression were not necessarily excluded from participation. The Institutional Review Board of Vanderbilt University approved all procedures. Written consent and assent was obtained from all children and participants.

Forty-one students were classified as right-handed, six as ambidextrous, and two as left-handed (Edinburgh Handedness Inventory; Oldfield, 1971). Two students had a confirmed diagnosis of ADHD and were regularly receiving medication at the time of the scan. No other behavioral disorders were confirmed. Scanning took place during the summer or school year after the completion of the grade that corresponds to the TCAP tests utilized for the analysis in this study. Descriptive and cognitive data and information regarding time between scan and TCAP are presented in Table 2. Although participants in the current sample demonstrated a large range in both TCAP math and TCAP reading scores, mean scores were at or above state averages for both measures. Raw scores on the TCAP math and TCAP reading were *z*-transformed using normative data obtained from the state of Tennessee Department of Education. Shapiro-Wilk tests show that neither measure violated normality (TCAP math $p = .0.25$, TCAP reading $p = .49$), suggesting that the present sample comprised a representative range of math performance.

2.2 | Standardized cognitive measures used for assessment

The Wechsler Abbreviated Scale of Intelligence (WASI), including the vocabulary, similarities, block design, and matrix reasoning subtests, was administered at the time of scan as a measure of global IQ (Weschler, 1999). The vocabulary and similarities subtests combine to form a measure of Verbal IQ. The block design and matrix reasoning subtests provide a measure of Performance IQ. Verbal IQ and Performance IQ are combined as a measure of Full Scale IQ (bivariate correlations Table 3).

2.3 | Grade-level achievement test

Student scores on the Tennessee Comprehensive Assessment Program (TCAP) were obtained from parents when they visited the lab and were administered at school during four school years from 2010 to 2013. It includes subtests in Reading/Language Arts (Reading), Mathematics, Science, and Social Studies. Only the Math and Reading subtests were used for this study. The TCAP is an end-of-year assessment, typically administered in April of each year during Tennessee's August–June school year. The math subtest has a duration of 84–93 minutes depending on grade level and year of test administration. It contains items in the following categories: (1) mathematical processes, (2) number and operations, (3) algebra, (4) geometry and measurement, and (5) data analysis, statistics, and probability. For example, in 3rd grade, items considered for the *number and operations* category would include identifying the place value of numbers in the ones, tens, hundreds, thousands, and ten-thousands positions. Items in the same category in 8th grade would include using scientific notation to compute products and quotients (see Supplementary Table S1 for more detailed description of content at each grade level).

TABLE 4 Descriptive statistics of grade-level split groups

	3rd & 4th grade (n = 25)	5th to 8th grade (n = 24)
Age at scan (years)	10.5 (*0.52*)	12.1 (*1.16*)
Grade	3.8 (*0.41*)	5.63 (*0.88*)
TCAP to scan (months)	8.2 (*4.00*)	8.2 (*3.12*)
TCAP math (z-score)	0.60 (*0.94*)	0.43 (*0.78*)
TCAP reading (z-score)	0.54 (*0.99*)	0.59 (*0.95*)
Performance IQ (WASI)*	106.1 (*14.94*)	105.0 (*11.91*)
Verbal IQ (WASI)	111.0 (*14.36*)	112.9 (*15.10*)
Full Scale IQ (WASI)*	109.8 (*14.02*)	110.0 (*13.48*)

Note. *SD*, standard deviation; TCAP, Tennessee Comprehensive Assessment Program; WASI, Wechsler Abbreviated Scale of Intelligence.
*n = 48 for Performance IQ and Full Scale IQ.

The reading subtest has a duration of 123–150 minutes depending on grade level and year of test administration. It contains items in the following categories: (1) Language, (2) Vocabulary, (3) Writing and Research, (4) Communication and Media, (5) Logic, (6) Informational Text, and (7) Literature. Items on the test are directly aligned with the state curriculum content standards (https://www.tn.gov/education/article/academic-standards-archive). Both math and reading scores are z-scores derived from an individual's score compared to means and standard deviations of students across all of Tennessee who took the same test, in the same year, at the same grade level as the individual represented in this sample.

2.4 | Image acquisition

T1-weighted MRI was performed on a Philips Achieve 3T scanner with an 8-channel head coil in the year following TCAP achievement tests. Average time in months between TCAP and scan was M = 8.2 and ranged from 1.5 to 13.5. Magnetization Prepared Rapid Gradient Recalled Echo (MP-RAGE; Mugler & Brookeman, 1990) anatomical scans were acquired according to the following parameters: 256 × 256 scan resolution; 170 slices; 1 mm slice thickness; 7.975s TR; 3.67s TE; Flip Angle = 7°; voxel size 1 mm^3 isotropic; acquisition time 392s; oriented AC-PC.

2.5 | Voxel-based morphometry

Images were analyzed using SPM8 (Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk), on a MATLAB platform (version 8.6, Mathworks, Natick, MA). Anatomical images for all analyses were processed according to the VBM protocol described by Ashburner (2010) with the following specifications. T1-weighted structural scans were first segmented to obtain separate GM, WM, and cerebral spinal fluid (CSF) images (Ashburner & Friston, 2005). GM, WM, and CSF values were summed to create a metric of global brain volume (GBV). Second, a population-specific template was created using diffeomorphic anatomical registration (DARTEL) (Ashburner, 2007). Third, each subject's GM map was transformed to the customized template space based on the sample of 49 participants and then normalized into MNI space by coregistering with the Montreal Neurological Institute (MNI152) brain template. The warped images were modulated by the Jacobian determinants derived from DARTEL to obtain maps of GM volume maintaining an isotropic voxel resolution of 1.0 × 1.0 × 1.0 mm^3. Two options exist in SPM8 for handling the effects of warping of GM that affect subsequent interpretation (Mechelli, Price, Friston, & Ashburner, 2005). One option is to leave voxel intensities 'unmodulated' thereby preserving the concentration of GM in each voxel and changing the total amount of GM. Analyses on unmodulated images should be interpreted as findings related to GM *concentration* or *density*. The second option is to scale the intensity of GM by the Jacobian determinants derived from spatial normalization at each voxel, 'modulated' normalization. This procedure results in preserved volumetric data. All subsequent analyses were performed on 'modulated' data utilizing the second option of normalization and thus the present results are interpreted in terms of regional GM volume. In the final stage of preprocessing, the GM volume maps were smoothed with a full-width at half-maximum Gaussian kernel of 10 mm to normalize the data.

2.6 | Grade-level math achievement and whole-brain grey matter volume

To investigate whether regional GM volume relates to individual differences in math achievement scores on the TCAP, we ran a linear regression model with TCAP math scores as the independent variable and VBM-derived GM volume maps as the dependent variable. Age at time of scan, sex, GBV, and TCAP reading were included as covariates in the model to control for anatomical differences related to age, sex, overall brain volume, and reading ability. All independent variables were orthogonalized with respect to the previous variables in the regression model. The regression analysis was run on the whole brain with a GM mask and an absolute threshold mask of 0.01. A cluster-level correction threshold of $p < .05$ was applied using the REST AlphaSim correction (uncorrected $p < .001$, minimum cluster threshold: 403 voxels) (Song et al., 2011, http://restfmri.net/forum/REST).

2.7 | Grade-level median split contrast

Previous research suggests that the neural correlates of math competence vary throughout development (Ranpura et al., 2013; Rosenberg-Lee, Barth, & Menon, 2011; Venkatraman, Ansari, & Chee, 2005). Therefore, to investigate whether the above relationship differs as a function of grade, we split the sample into two groups with a median grade-level split and ran a contrast that compared the slope of relationship between GM volume and TCAP math scores of the lower grade levels to the slope of the higher grade levels. Only sex and TCAP reading were included as covariates of non-interest. The split resulted in one group of students in 3rd and 4th grade (n = 25) and one group of students in 5th to 8th grade (n = 24) (Table 4). These groups did not differ in performance IQ, verbal IQ, TCAP reading, TCAP math, or time

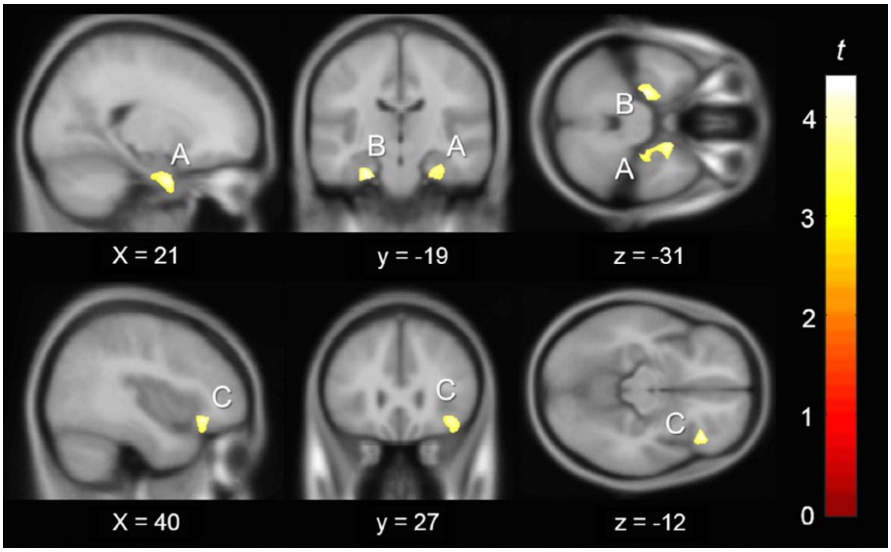

FIGURE 1 Significant clusters of GM related to TCAP math after controlling for GBV, age, sex, and TCAP reading ($p < .05$ cluster corrected, $p < .001$ uncorrected) overlaid on the MNI 152 template ($n = 49$). MNI slice coordinate listed. Cluster letters correspond to Table 5. (A) R Hippocampal Formation (B) L Hippocampal Formation (C) R inferior frontal gyrus, *pars orbitalis*

between scan and TCAP ($p > .48$ for all t-tests) or in variance of the same variables ($p > .25$ for all Levene's tests). Analysis was run using a GM mask and an absolute threshold mask of 0.01. A cluster-level correction threshold of $p < .05$ was applied (uncorrected $p < .001$, minimum cluster threshold: 403 voxels) as in the original analysis.

3 | RESULTS

3.1 | Grade-level math achievement and whole-brain grey matter volume

This analysis revealed three clusters of voxels in which greater GM volume was related to higher TCAP math scores after controlling for age, sex, global brain volume, and TCAP reading scores. No clusters showing the inverse relationship were found. The three clusters were identified as the large portions of the left and right hippocampal formations including the hippocampus proper (CA1, CA3, Dentate Gyrus, Entorhinal Cortex, and Subiculum) and the right inferior frontal gyrus, *pars orbitalis* (right IFG*orb*) (Figure 1, Table 5). Cohen's d effect sizes were calculated using the formula $d = \frac{2t}{\sqrt{df}}$ from peak t-values from each cluster ($df = 44$). Anatomical descriptions of each cluster were identified from maximum probabilities derived from SPM 8 Anatomy Toolbox v2.15 (Eickhoff et al., 2005) (Table 5). Due to the relatively high variability in time between TCAP and scan (Table 2), the regression was repeated with the inclusion of time between TCAP and scan as a covariate. Results again revealed three clusters of voxels of the same anatomical description as the original analysis showing a positive relationship between GM volume and higher TCAP math. In other words, inclusion of time between TCAP and scan as a covariate did not alter the pattern of results. Because TCAP math and TCAP reading are highly correlated ($r = .758$), we calculated variance inflation factors (VIFs) for each covariate in the regression model. All covariates had VIFs of less than 2.5 (TCAP reading = 2.40), indicating low inflation due to collinearity (Menard, 1995). To investigate the relationship between grey matter and math independent of the influence of reading, we conducted an additional analysis in which reading was not included as a covariate, continuing to control for overall global brain volume, age, and sex. One large cluster in the right middle temporal gyrus showed a positive relationship between GMV and scores on TCAP math ($p < .05$ cluster corrected, $p < .001$ uncorrected, voxels = 833, see Supplementary Figure S1 for details). The relation between grey matter volume in left hippocampal formation, right hippocampal formation, and right IFGorb region and math competence remained significant only at an uncorrected threshold of $p < .05$, uncorrected.

3.2 | Grade-level median split contrast

The contrast between the slopes of the 3rd–4th grade group and the 5th–8th grade group revealed three significant clusters (Figure 2, Table 6). One negative cluster was identified in the posterior portion of the left AG (PGp). This means that in the 3rd–4th grade group, GM volume in the left AG was more strongly related to TCAP math scores than it was for 5th–8th grade students. In order to ensure that the left AG was related to TCAP math for this group, independent of its difference from the 5th–8th grade group, we ran a whole-brain regression analysis predicting GM from TCAP math, controlling for sex and TCAP reading, which included only the 3rd–4th grade group. The same left AG cluster was positively correlated with TCAP math at $p < .005$, uncorrected (voxels = 1043). Conversely, two positive clusters were identified, one in the left inferior frontal gyrus, *pars orbitalis* (left IFG*orb*) and one in the left cerebellum, *lobules VIIa & VIIIb* (Diedrichsen, Balsters, Flavell, Cussans, & Ramnani, 2009), where the relationship between GM volume and TCAP math scores was stronger for the 5th–8th grade students. In order to ensure that the left IFG*orb* and left cerebellum were related to TCAP math for this group, independent of its difference from the 3rd–4th grade group, we ran a whole-brain regression analysis predicting GM from TCAP math, controlling for sex and TCAP reading, which included only the 5th–8th grade group. The left IFG*orb* was positively correlated with TCAP math at $p < .005$, uncorrected (voxels = 126); however, the left cerebellum cluster was not.

TABLE 5 VBM cluster results for grey matter analysis

Cluster	Peak MNI (x y z)	Cluster Size	Peak *t*	Effect Size**	BA	Anatomical description
A	(19 2 −34)	2866	4.39	1.32	27, 28, 34, 35	R Hippocampal Formation
B	(−22 −14 −34)	1481	4.25	1.30	27, 28, 34, 35	L Hippocampal Formation
C	(39 29 −12)	913	4.16	1.25	47	R Inferior Frontal Gyrus, *pars orbitalis*

*All results cluster corrected at $p < .05$, voxels = 403, uncorrected $p < .001$.
**All effect sizes reported as Cohen's *d*.

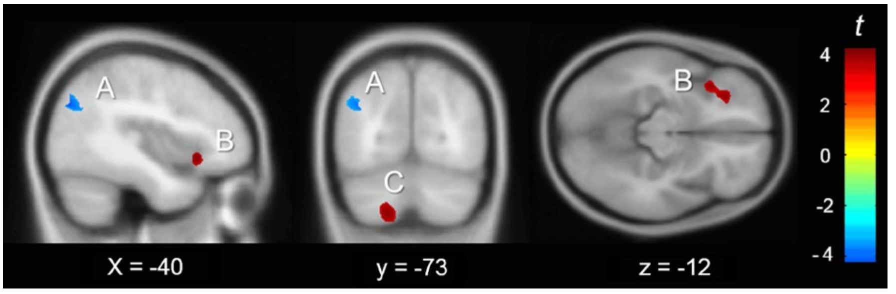

FIGURE 2 Significant clusters where slope of TCAP math predicting GMV is significantly different for grade-level split. Positive t-values indicate a greater slope for the higher grade levels (5th to 8th), negative *t*-values indicate a greater slope for the lower grade levels (3rd and 4th). Overlaid on the MNI 152 template ($n = 49$). MNI slice coordinate listed. Cluster letters correspond to Table 6. (A) L Angular Gyrus, *PGp* (B) L IFG, *orb* (C) L Cerebellum, *lobules VIIa & VIIIb*

TABLE 6 VBM cluster results for grade-level split

Cluster	Peak MNI (x y z)	Cluster Size	Peak *t*	Effect Size**	BA	Anatomical description
A	(−40 −76 32)	741	4.17	1.29	39	L Angular Gyrus, *PGp*
B	(−29 32 −10)	893	4.16	1.28	47	L inferior frontal gyrus, *pars orbitalis*
C	(−17 −72 −50)	1016	3.86	1.19	–	L Cerebellum, *lobules VIIa* & *VIIIb*

*All results cluster corrected at $p < .05$, voxels = 403, uncorrected $p < .001$.
**All effect sizes reported as Cohen's *d*.

4 | DISCUSSION

The current study is the first to investigate the relationship between regional differences in GM volume and performance in grade-level mathematics as measured by a state-wide, school-based test of math achievement. We observed three regions in which greater regional GM volume was related to better performance on the TCAP math subtest when controlling for global brain volume, age, sex, and TCAP reading. These regions include the right and left hippocampal formations (including the hippocampus proper, entorhinal cortex, and subiculum) and the right inferior frontal gyrus (IFG*orb*). Effect sizes were strongest in the hippocampal formations bilaterally and peaked in the entorhinal cortex of each cluster (right $d = 1.32$, left $d = 1.30$).

In contrast with previous literature linking structural integrity and functional activation patterns of the IPS with math achievement (Han et al., 2008; Isaacs et al., 2001; Lubin et al., 2013; Price et al., 2016; Ranpura et al., 2013; Rotzer et al., 2008), GM volume of the IPS was not related to TCAP math performance in the current study. The three regions observed in the present study are associated with a wide range of domain-general cognitive abilities. The bilateral hippocampal formation has been consistently associated with generalized learning behaviors such as the encoding and retrieval of declarative memories (Eichenbaum, 2000) including but not limited to arithmetic fact knowledge (Cho et al., 2012; Qin et al., 2014). The IFG*orb* is involved in sensory integration, affective value reinforcers, decision-making, and visuo-spatial working memory (Kringelbach & Radcliffe, 2005; Metcalfe, Ashkenazi, Rosenberg-Lee, & Menon, 2013; Rotzer et al., 2009). In addition to their domain-general roles, the hippocampus and IFG*orb* have been previously associated with math development. Cho et al. (2012) demonstrated that children with higher arithmetic retrieval fluency have greater activation in the right hippocampus and parahippocampal gyrus during arithmetic problem solving, while Supekar et al. (2013) showed that increased GM volume in the hippocampus was related to gains in arithmetic ability in response to a tutoring intervention. Starke et al. (2013) reported an association between white matter in the right IFG and the numerical distance effect in 6–7-year-old children, and Rotzer et al. (2008) reported that children with Dyscalculia have decreased GM in the left IFG. Thus, in concert with previous findings, the present results suggest that the domain-general functions supported by the hippocampus and IFG play a crucial role in the effective acquisition of school-relevant math skills.

While domain-specific mechanisms such as numerical magnitude processing are undoubtedly important factors in the development of mathematical knowledge, the current findings suggest that school-based math competence is associated primarily with higher GM volume in regions known to have a domain-general role in learning (i.e., bilateral hippocampus and the right IFG*orb*). It should be noted that there were substantial differences between regression models with and without the inclusion of TCAP reading as a covariate. However, given that appropriate statistical procedures were undertaken to avoid multicollinearity and that the metrics for inflation due to multicollinearity are low, it is unlikely that the current findings are a spurious result of multi-collinearity. Instead, we suggest that the differences in results when controlling for versus not controlling for TCAP reading likely reflect meaningful differences between the isolating academic processes and associated neural correlates specific to math by controlling for another, broad measure of academic achievement, reading, versus not. An alternative explanation of differences between the two models is that individual differences *between* reading and math scores (i.e., discrepancy) within an individual are influencing the models, but it is difficult to conceptualize why a discrepancy between the measures would be associated with a specific neural correlate, and thus this explanation seems unparsimonious. Future research should confirm the nature of this relationship by exploring this relationship with the inclusion of other academic measures.

To investigate the potential influence of grade level on our findings, we contrasted the relation between GM volume and TCAP math scores in grades 3–4 versus grades 5–8. Students in grades 3–4 showed a stronger positive relationship between GM and TCAP math in the left AG than students in the higher grades. The left AG has been associated with numerical symbol processing and individual differences in arithmetic fluency retrieval (Delazer et al., 2005; Grabner et al., 2007; Price & Ansari, 2011) and thus, the present results suggest these processes are more important for math performance in the lower grade levels. Although results at the whole-group level do not support our domain-specific hypothesis, results from the grade-level split may indicate the importance of this relationship earlier in development. Conversely, students in grades 5–8 showed a stronger positive relationship between GM volume and TCAP math in the left IFG*orb* and the cerebellum. Similar to the right IFG*orb* found in the whole-group analysis, the left IFG*orb* has been implicated in rule-learning, decision-making, and visuospatial working memory, suggesting that math skills in the higher grade levels are drawing on a greater variety of domain-general executive functions. Indeed, TCAP content in the lower grades tends to focus on basic computations of numerical quantities, while in the higher grades it requires the utilization of strategies, orders of operations, and rule-learning. The left cerebellum region was not present in the whole-brain analysis or either of the independent analyses with the 3rd–4th or 5th–8th grade groups and thus further empirical investigation is required before we can draw conclusions on the mechanisms underlying this result. Further, applying a grade-level split was necessary due to unequal representation of grade level in our sample, which results in a rather coarse comparison of the mean correlations between grey matter and math in each group. It would be interesting in future studies to ensure equal distributions of participants among grades in order to gain a more nuanced understanding of neuroanatomical changes across year of schooling.

It is possible that the observed grade-level differences could be due to neuroanatomical maturation as opposed to changes in content on the TCAP itself. However, given that the AG is associated with fluency of arithmetic fact retrieval and arithmetic proficiency in adults (Delazer et al., 2005; Grabner et al., 2007), it seems more likely that the current results are driven by content differences in the TCAP across grades. Nonetheless, further empirical investigation is required to confirm this speculation.

In conclusion, the current results suggest two significant implications for our understanding of the neural correlates of math competence and the future of educational neuroscience research, respectively. First, the neuroanatomical mechanisms most related to math competence measured by in-school math tests differ from those related to performance on standardized math measures. The former are related to domain-general cognitive mechanisms associated with memory encoding and retrieval, while the latter are associated with domain-specific numerical processing regions (Li et al., 2013; Price et al., 2016). This suggests that when investigating math competence, standardized measures may not always be an appropriate proxy for the skills and abilities necessary for in-school math achievement. Furthermore, the neural correlates most related to math competence depend on the math content being assessed, and thus, considering the role of development and environment in the relation between neuroscience and education is of critical importance.

The second and related implication stemming from these results is that for neuroscience to bridge the gap to education, future research should utilize math measures closely related to the learning environment in addition to more traditional standardized measures of math achievement. Although the current results do not offer any direct suggestions for the improvement of educational approaches, they offer insights distinct from previous literature focused on standardized achievement measures. The two types of performance measures may provide unique and complementary insights into the neural correlates of math learning that pave the way for more effective pedagogical approaches and improved educational outcomes.

ACKNOWLEDGEMENTS

This work was supported by RO1 HD044073, P30 HD015052, RO1 HD067254, U54 HD083211 and CTSA RR024975.

REFERENCES

Aboud, K.S., Bailey, S.K., Petrill, S.A., & Cutting, L.E. (2016). Comprehending text versus reading words in young readers with varying reading ability: Distinct patterns of functional connectivity from common processing hubs. *Developmental Science*, *19*, 632–656.

Ansari, D., & Dhital, B. (2006). Age-related changes in the activation of the intraparietal sulcus during nonsymbolic magnitude processing: An event-related functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *18*, 1820–1828.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*, 95–113.

Ashburner, J. (2010). VBM tutorial. Retrieved from http://www.fil.ion.ucl.ac.uk/~john/misc/VBMclass10.pdf

Ashburner, J., & Friston, K.J. (2005). Unified segmentation. *NeuroImage*, *26*, 839–851.

British Council (2010). STEM Education Programme. Retrieved 1 September 2015, from http://www.britishcouncil.org/education/science/newton/stem-education-programme

Bruer, J.T. (1997). Education and the brain: A bridge too far. *Educational Researcher*, *26*, 4–16.

Bruer, J.T. (2006). Points of view: On the implications of neuroscience research for science teaching and learning: Are there any? *CBE Life Sciences Education*, *5*, 110–110.

Bugden, S., & Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition*, *118*, 32–44.

Bugden, S., Price, G.R., McLean, D.A., & Ansari, D. (2012). The role of the left intraparietal sulcus in the relationship between symbolic number processing and children's arithmetic competence. *Developmental Cognitive Neuroscience*, *2*, 448–457.

Bulthé, J., De Smedt, B., & Op de Beeck, H.P. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *NeuroImage*, *87*, 311–322.

Bynner, J., & Parsons, S. (1997). *Does numeracy matter?* London: The Basic Skills Agency.

Bynner, J., & Parsons, S. (2006). *Does numeracy matter more?* London: National Research and Development Centre for Adult Literacy and Numeracy.

Cantlon, J.F., & Brannon, E.M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science*, *17*, 401–406.

Cantlon, J.F., Brannon, E.M., Carter, E.J., & Pelphrey, K.A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology*, *4*, 844–854.

Cho, S., Metcalfe, A.W.S., Young, C.B., Ryali, S., Geary, D.C., & Menon, V. (2012). Hippocampal–prefrontal engagement and dynamic causal interactions in the maturation of children's fact retrieval. *Journal of Cognitive Neuroscience*, *9*, 1849–1866.

Delazer, M., Ischebeck, A., Domahs, F., Zamarian, L., Koppelstaetter, F., Siedentopf, C.M., … & Felber, S. (2005). Learning by strategies and learning by drill: Evidence from an fMRI study. *NeuroImage*, *25*, 838–849.

Diedrichsen, J., Balsters, J.H., Flavell, J., Cussans, E., & Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *NeuroImage*, *46*, 39–46.

Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., … Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446.

Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nature Reviews Neuroscience*, *1*, 41–50.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*, 1325–1335.

Emerson, R.W., & Cantlon, J.F. (2012). Early math achievement and functional connectivity in the fronto-parietal network. *Developmental Cognitive Neuroscience*, *2*(Suppl. 1), S139–S151.

Grabner, R.H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., & Neuper, C. (2007). Individual differences in mathematical competence predict parietal brain activation during mental calculation. *NeuroImage*, *38*, 346–356.

Halberda, J., Mazzocco, M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665–668.

Han, Z., Fuchs, L., Davis, N., Cannistraci, C.J., Anderson, A.W., Gore, J.C., & Dawant, B.M. (2008). Analysis of anatomic variability in children with low mathematical skills. *Proceedings of SPIE*, *6916*, 69160S–69160S-8.

Hubbard, E.M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, *6*, 435–448.

Isaacs, E.B., Edmonds, C.J., Lucas, A., & Gadian, D.G. (2001). Calculation difficulties in children of very low birthweight: A neural correlate. *Brain: A Journal of Neurology*, *124*, 1701–1707.

Kringelbach, M.L., & Radcliffe, J. (2005). The human orbitofrontal cortex: Linking reward to hedonic experience. *Nature Reviews Neuroscience*, *6*, 691–702.

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., … & von Aster, M. (2011). Mental number line training in children with developmental dyscalculia. *NeuroImage*, *57*, 782–795.

Li, Y., Hu, Y., Wang, Y., Weng, J., & Chen, F. (2013). Individual structural differences in left inferior parietal area are associated with schoolchildrens' arithmetic scores. *Frontiers in Human Neuroscience*, *7*, 844.

Lubin, A., Rossi, S., Simon, G., Lanoë, C., Leroux, G., Poirel, N., … & Houdé, O. (2013). Numerical transcoding proficiency in 10-year-old schoolchildren is associated with gray matter inter-individual differences: A voxel-based morphometry study. *Frontiers in Psychology*, *4*, 197.

Mazzocco, M.M.M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development*, *82*, 1224–1237.

Mechelli, A., Price, C.J., Friston, K.J., & Ashburner, J. (2005). Voxel-based morphometry of the human brain: Methods and applications. *Current Medical Imaging Reviews*, *1*, 105–113.

Menard, S. (1995). *Applied logistic regression analysis: Sage university series on quantitative applications in the social sciences*. Thousand Oaks, CA: Sage.

Metcalfe, A.W.S., Ashkenazi, S., Rosenberg-Lee, M., & Menon, V. (2013). Fractionating the neural correlates of individual working memory components underlying arithmetic problem solving skills in children. *Developmental Cognitive Neuroscience*, *6*, 162–175.

Mugler, J.P., & Brookeman, J.R. (1990). Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magnetic Resonance in Medicine*, *15*, 152–157.

OECD (2010). *The high cost of low educational performance: The long-run economic impact of improving Pisa outcomes*. Paris: OECD.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Price, G.R., & Ansari, D. (2011). Symbol processing in the left angular gyrus: Evidence from passive perception of digits. *NeuroImage*, *57*, 1205–1211.

Price, G.R., Holloway, I., Räsänen, P., Vesterinen, M., & Ansari, D. (2007). Impaired parietal magnitude processing in developmental dyscalculia. *Current Biology*, *17*, R1042–R1043.

Price, G.R., Wilkey, E.D., Yeo, D.J., & Cutting, L.E. (2016). The relation between 1st grade grey matter volume and 2nd grade math competence. *NeuroImage*, *124*, 232–237.

Qin, S., Cho, S., Chen, T., Rosenberg-Lee, M., Geary, D.C., & Menon, V. (2014). Hippocampal-neocortical functional reorganization underlies children's cognitive development. *Nature Neuroscience*, *17*, 1263–1269.

Ranpura, A., Isaacs, E., Edmonds, C., Rogers, M., Lanigan, J., Singhal, A., … Butterworth, B. (2013). Developmental trajectories of grey and white matter in dyscalculia. *Trends in Neuroscience and Education*, *2*, 56–64.

Reich, W., Leacock, N., & Shanfeld, K. (1997). DICA-IV Diagnostic Interview for children and Adolescents-IV.

Ritchie, S.J., & Bates, T.C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science*, *24*, 1301–1308.

Rosenberg-Lee, M., Ashkenazi, S., Chen, T., Young, C.B., Geary, D.C., & Menon, V. (2015). Brain hyper-connectivity and operation-specific deficits during arithmetic problem solving in children with developmental dyscalculia. *Developmental Science*, *18*, 351–372.

Rosenberg-Lee, M., Barth, M., & Menon, V. (2011). What difference does a year of schooling make? Maturation of brain response and connectivity between 2nd and 3rd grades during arithmetic problem solving. *NeuroImage*, *57*, 796–808.

Rotzer, S., Kucian, K., Martin, E., von Aster, M., Klaver, P., & Loenneker, T. (2008). Optimized voxel-based morphometry in children with developmental dyscalculia. *NeuroImage*, *39*, 417–422.

Rotzer, S., Loenneker, T., Kucian, K., Martin, E., Klaver, P., & von Aster, M. (2009). Dysfunctional neural network of spatial working memory contributes to developmental dyscalculia. *Neuropsychologia*, *47*, 2859–2865.

Rykhlevskaia, E., Uddin, L.Q., Kondos, L., & Menon, V. (2009). Neuroanatomical correlates of developmental dyscalculia: Combined evidence from morphometry and tractography. *Frontiers in Human Neuroscience*, *3*, 51.

Song, X.-W., Dong, Z.-Y., Long, X.-Y., Li, S.-F., Zuo, X.-N., Zhu, C.-Z., … & Zang, Y.-F. (2011). REST: A toolkit for resting-state functional magnetic resonance imaging data processing. *PLoS ONE*, *6*, e25031.

Starke, M., Kiechl-Kohlendorfer, U., Kucian, K., Pupp Peglow, U., Kremser, C., Schocke, M., & Kaufmann, L. (2013). Brain structure, number magnitude processing, and math proficiency in 6- to 7-year-old children born prematurely: A voxel-based morphometry study. *NeuroReport*, *24*, 419–424.

Supekar, K., Swigart, A.G., Tenison, C., Jolles, D.D., Rosenberg-Lee, M., Fuchs, L., & Menon, V. (2013). Neural predictors of individual differences in response to math tutoring in primary-grade school children. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 8230–8235.

Venkatraman, V., Ansari, D., & Chee, M.W.L. (2005). Neural correlates of symbolic and non-symbolic arithmetic. *Neuropsychologia*, *43*, 744–753.

Weschler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.

White House Office of the Press Secretary (2010). President Obama expands 'Educate to Innovate' Campaign for Excellence in Science, Technology, Engineering, and Mathematics (STEM) [Press Release] Education. Retrieved from https://www.whitehouse.gov/the-press-office/president-obama-expands-educate-innovate-campaign-excellence-science-technology-eng

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Wilkey ED, Cutting LE, and Price GR. Neuroanatomical correlates of performance in a state-wide test of math achievement. *Dev Sci*. 2017;00:e12545. https://doi.org/10.1111/desc.12545