Developmental Cognitive Neuroscience 46 (2020) 100873

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

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

Meaning to multiply: Electrophysiological evidence that children and adults treat multiplication facts differently

Amandine E. Grenier [a,*], Danielle S. Dickson [b], Corey S. Sparks [c], Nicole Y.Y. Wicha [a,d,e]

[a] *Department of Biology, The University of Texas at San Antonio, San Antonio, TX, USA*
[b] *Department of Psychology and Center for Language Science, The Pennsylvania State University, Pennsylvania, USA*
[c] *Department of Demography, The University of Texas at San Antonio, TX, USA*
[d] *UTSA Neurosciences Institute, The University of Texas at San Antonio, San Antonio, TX, USA*
[e] *Research Imaging Institute, UT Health San Antonio, San Antonio, USA*

A R T I C L E I N F O

Keywords:
N400
Multiplication
Arithmetic
Language
Children
Adults

A B S T R A C T

Multiplication tables are typically memorized verbally, with fluent retrieval leading to better performance in advanced math. Arithmetic development is characterized by strategy shifts from procedural operations to direct fact retrieval, which would not necessitate access to the facts' conceptual meaning. This study tested this hypothesis using a combination of event related brain potentials (ERP) and behavioral measures with 3rd-5th grade children and young adults. Participants verified the solutions to simple multiplication problems ($2 \times 3 = 6$ or $= 7$) and the semantic fit of word-picture pairs, separately. Children showed an N400 effect to multiplication solutions with larger (more negative) amplitude for incorrect than correct solutions, reflecting meaning-level processing. A similar ERP response was observed in the word-picture verification task, with larger negative amplitude for word-picture pairs that were semantically mismatched compared to matched. In contrast, adults showed a P300 response for *correct* solutions, suggesting that they treated these solutions as potential targets in over-rehearsed mathematical expressions. This P300 response was specific to math fact processing, as the word-picture verification task elicited a classic N400 in adults. These ERP findings reveal an overlooked developmental transition that occurs after fifth grade, and speak to theories of arithmetic that have been based primarily on adult data.

1. Introduction

Arithmetic facts, like multiplication tables, may be linked to language given that children memorize them through verbal rehearsal (Siegler, 1988). Neuroimaging studies of exact arithmetic in children support this, revealing a developmental shift from relying more on parietal areas in the brain, associated with estimation and calculation, to temporal and frontal regions, associated with verbal retrieval (Kawashima et al., 2004; Peters and De Smedt, 2018; Prado et al., 2014). Behavioral findings suggest that this shift to retrieval may happen early in learning multiplication facts (Geary, 2011; Jordan et al., 2003). Similarly, a small number of event-related potential (ERP) studies have suggested that children exhibit an adult-like brain response early on, in-line with quickly adopting a retrieval strategy (Prieto-Corona et al., 2010; Xuan et al., 2007; Zhou et al., 2011). However, some behavioral researchers have argued that the developmental transition between calculation and retrieval may happen more gradually (LeFevre et al., 1996; Siegler, 1996). Moreover, recent neuroimaging studies with adults have reinterpreted the ERP response observed to mathematical fact expressions, implying that children might not process arithmetic like adults do (Dickson et al., 2018; Dickson and Federmeier, 2017; Jasinski and Coch, 2012). The current study adds critical new evidence to the limited literature on the neurocognitive basis of arithmetic in elementary school children and tests whether a slower developmental trajectory toward adult-like processing has indeed been overlooked.

Research using ERPs to study how arithmetic is processed has revealed robust brain responses to proposed solutions in both children and adults. This research has been dominated by studies that include adult populations (Dickson et al., 2018; Dickson and Federmeier, 2017; Jasinski and Coch, 2012; Jost et al., 2003; Martinez-Lincoln et al., 2015; Niedeggen et al., 1999; Núñez-Peña et al., 2006; Salillas and Wicha, 2012), with a smaller group of studies that have compared child and

* Corresponding author.
E-mail address: amandine.grenier@utsa.edu (A.E. Grenier).

https://doi.org/10.1016/j.dcn.2020.100873
Received 19 November 2019; Received in revised form 13 October 2020; Accepted 16 October 2020
Available online 20 October 2020

adult brain responses (Moore et al., 2014; Prieto-Corona et al., 2010; Xuan et al., 2007; Zhou et al., 2011). The methods used in these studies have varied, with some using verification tasks and others production, or different operations (addition vs multiplication, etc.). In turn, the brain responses elicited by these paradigms also varied. However, the studies have generally reported the same underlying ERP components for adults and children, with only differences in where (scalp distribution) or how large (amplitude) the effect was between the groups. The interpretation of the results was never that children and adults showed qualitatively different brain responses. For example, Prieto-Corona et al. (2010) reported a larger negative-going amplitude for incorrect than correct multiplication solutions (e.g., 2 × 4 = 9 vs = 8) in both children and adults, with a larger effect (larger amplitude difference between correct and incorrect) for children. This ERP effect was interpreted as a modulation of a commonly reported ERP component, the N400, in both groups. The N400 is a component that emerges in response to any potentially meaningful stimulus, like a word or a picture, and decreases in amplitude if the stimulus is contextually supported, for example, if a word can be anticipated based on preceding sentence context (Kutas and Federmeier, 2011; Kutas and Hillyard, 1980; Wicha et al., 2003). As such, it has been argued that this N400 effect in children reflects meaning-level processing when judging the correctness of arithmetic facts (Cerda et al., 2019).

Seminal ERP studies with adults had similarly argued that arithmetic expressions elicit brain responses analogous to the N400 observed for words that vary in expectancy within a sentence context (Jost et al., 2003; Niedeggen et al., 1999). Initial research therefore implied that both children and adults construct meaning-level representations for simple multiplication problems. However, as alluded to above, more recent studies from multiple labs have reexamined the nature of this arithmetic correctness effect in adults and came to the same independent conclusion that the adult brain response was functionally and morphologically different than the N400 (Dickson et al., 2018; Dickson and Federmeier, 2017; Dickson and Wicha, 2019; Jasinski and Coch, 2012; Wicha et al., 2018). Essentially, what was originally interpreted as an N400 modulation, with larger amplitude for incorrect solutions, was more in-line with a modulation of the P300, with larger amplitude for correct solutions (see Dickson and Wicha, 2019 for a more in-depth discussion of this reexamination of the adult correctness effect). A P300 is typically observed during stimulus categorization and is larger for task-relevant targets than non-target items (Polich, 1987, 2007, 2012; Sutton et al., 1965, 1982). Studies that measured the pattern of ERPs in adults across multiple methodological manipulations have confirmed that the brain response to arithmetic facts is more consistent with the P300 than the N400 (Dickson and Wicha, 2019; Dickson et al., 2018). Thus, to the extent that the reports of the arithmetic N400 effect in children are true, this reinterpretation of the adult ERPs implies that children and adults may actually engage qualitatively different neurocognitive processes when verifying arithmetic facts, with only children reliably building meaning-level expectations for the solutions.

The N400 response observed in arithmetic fact verification also raises a potential conflict for current theories of mathematical cognition, which have been based primarily on behavioral and brain localization findings from adults (Campbell and Clark, 1988; Dehaene, 1992; McCloskey et al., 1992). These models do not explicitly speak to the N400 congruency effects reported in the literature. However, given that the N400 is thought to reflect access to meaning in memory, it may shed light on assumptions that these models make. Because memorized arithmetic facts are highly automated, it has been hypothesized that they do not require access to the conceptual understanding of the facts, namely their numerosity (Cipolotti et al., 1991; Dehaene and Cohen, 1997; Naccache and Dehaene, 2001). However, decades of research have demonstrated that the N400 is elicited during attempted access to semantic memory for any potentially meaningful stimulus, whether it be a picture, word or other item (see Kutas and Federmeier, 2011 for review). Therefore, the N400 effect in children would suggest that their memory retrieval of arithmetic facts is not void of meaning, as implied by these models.

In the current study we directly examine the neurocognitive processes that children and adults engage using a multiplication verification task, and compare their brain and behavioral responses to a language task known to generate an N400 response. The goal of the study is to determine if children and adults engage similar neurocognitive processes when retrieving and verifying learned arithmetic facts, as has been suggested to date, or if a developmental difference in processing has been overlooked in the literature.

2. Methods

2.1. Participants

Experiment 1a. Child participants' (n = 99; 46 female) average age was 9 years 10 months (range 8 yrs 1 m - 11 yrs 9 m) and average grade was 4.6 (range 3.0–5.9). All were right-handed (average 0.80, range 0.11–1.00; abridged Edinburgh Inventory, Oldfield, 1971) with normal hearing and normal or corrected-to-normal vision. None had a history of developmental disorders nor took psychoactive medications. All participants completed offline measures prior to the EEG recording (Table 1). A subset (n = 62) returned for a second EEG session on average 4.5 months after the first (range: 2 days – 16 months), to complete the word-picture verification task. The returning participants' (30 female) average age was 10 years 6 months (range 8 yrs 8 m – 12 yrs) and average grade was 5.1 (range 3.2–5.9). Forty children moved up one grade level between session 1 (math verification task) and session 2 (word-picture task).

The sample was representative of the local community (San Antonio, Texas, USA) with diverse socioeconomic status (SES) (range 16.5–66; average 48.4 – considered upper-middle class; Hollingshead, 2011) and language backgrounds. Language profiles were established using an adapted LEAP-Q questionnaire (Marian et al., 2007). Participants included 65 monolinguals (English), as well as 14 English-dominant and 20 proficient Spanish-English bilinguals. All children provided verbal assent, and legal guardians provided written consent in accordance with the Institutional Review Board (IRB) of the University of Texas at San Antonio (UTSA).

Experiment 1b. Adult participants (n = 60, 35 female) were a representative sample of UTSA students (mean age 21.0 years, range 18.1 – 26.6 years), and included 32 monolingual English speakers and 28 fluent Spanish-English bilinguals (assessed via standardized tests and the language questionnaire). All were right-handed (average 0.80, range 0.50–1.00; abridged Edinburgh Inventory), had normal or corrected-to-normal vision and no cognitive or neurological disorders or current use of psychoactive substances. All participants consented in accordance with the UTSA IRB. Adults completed all tasks in a single session, with

Table 1
Offline performance on standardized cognitive measures.

	Children (n = 99)		Adults (n = 60)	
Assessments	Mean^ (SE)	Range	Mean (SE)	Range
Math fluency[1]				
Addition	101.60 (1.20)	71–142	96.27 (1.78)	74–123
Subtraction	103.69 (1.25)	76–152	98.95 (1.75)	75–122
Multiplication	106.91 (1.29)	86–160	97.23 (1.85)	65–117
Working memory[2]	106.05 (1.42)	74–149	101.37 (1.80)	72–133
Phonological awareness[3]	100.77 (1.45)	68–134	100.42 (1.92)	66–130

^ Standardized scores where 100 is the age-based norm and 15 points reflects 1 *SD* outside the norm.

[1] Math fluency test of the WIAT III (Wechsler, 2009).

[2] Numbers reversed test of the Woodcock-Johnson III Tests of Cognitive Abilities (Woodcock et al., 2001).

[3] Incomplete words test of the WJ-III Tests of Cognitive Abilities (Woodcock et al., 2001).

the word-picture verification task last.

2.2. Offline cognitive assessments

The same cognitive assessments were administered with children and adults. Arithmetic fluency was measured with 1-minute subtests for addition, subtraction and multiplication of the Wechsler Individual Achievement Test (WIAT III; Wechsler, 2009). To ensure minimum competence on the verification task, children had to meet a 3rd grade level for multiplication (sample range 9–40 out of 40 problems). Additional cognitive measures were collected before the EEG recording to facilitate comparison with other studies and are reported for completeness in Table 1.

2.3. Stimuli and procedure

The stimuli, procedures and EEG recording parameters were identical across both groups (children and adults) with the exception of the duration of the response period, as noted below. All experimental stimuli were presented using Paradigm (Perception Research Systems, 2007) on a 19-inch LCD monitor positioned 100 cm away from the participant. Auditory stimuli were presented through ER-1 insert headphones (Etymotic Research, Inc.). Participants judged the correctness of solutions and picture-matches by pressing one of two buttons on a LogiTech F310 Gamepad with their left and right index fingers; answer-response mapping was reversed for half the participants.

Multiplication verification task. All single-digit multiplication problems were used except those with operands 0 and 1 or ties (e.g., 5 × 5), which have been found to create atypical behavioral effects (see Campbell and Graham, 1985). The 56 problems were presented twice each, ending once with the correct solution and once with an incorrect solution, for a total of 112 trials. Incorrect solutions were table-related to the problem and generated as 1 or 2 plus or minus one of the operands (e.g., correct problem 2 × 4 = 8; add 1 to the first operand 3 x 4 = 12; incorrect table-related solution 2 × 4 = 12).

Multiplication problems appeared sequentially in the center of a computer monitor as Arabic numerals with no symbols (4 3 12). Participants were instructed to judge if the third number was the correct or incorrect product of the first two and respond as quickly as possible without sacrificing accuracy. The experiment was framed with a motivating child-friendly cover story about completing a mission on a rocket ship with the goal of collecting as many coins as possible. One coin was earned for each correctly answered multiplication problem. The number of coins collected opened different levels on a treasure box of rewards after the experiment. All children received a reward regardless of their performance. Feedback was given about how many coins had been earned after each block, but not on performance on individual trials.

Each experimental trial was structured as in Fig. 1. Participants were asked to respond as soon as possible after the solution appeared on the screen; most responses were recorded during the blank-screen period. The next trial began 1000 ms after registering a response, or after a maximum response period (5000 ms for children; 3000 ms for adults). Participants were asked to keep their eyes focused on the center of the screen to avoid eye and head movement as much as possible. Trials were presented in 8 blocks of 14 problems with self-paced breaks between blocks. Trials were pseudorandomized and fixed for all participants, so that no three incorrect or correct problems were presented in a row, and the same problem or solution never appeared consecutively.

Word-picture verification task. The purpose of this task was to generate an N400 in children for comparison to the brain response in the multiplication verification task. We used a word-picture verification task, which generates a robust N400 in this age group (e.g., Friedrich and Friederici, 2004). Line drawings of imageable nouns (objects or animals) from Snodgrass and Vanderwart (1980) were inverted as white drawings on black backgrounds (to reduce eye strain from the brightness of a white background). The nouns were the most common names for each

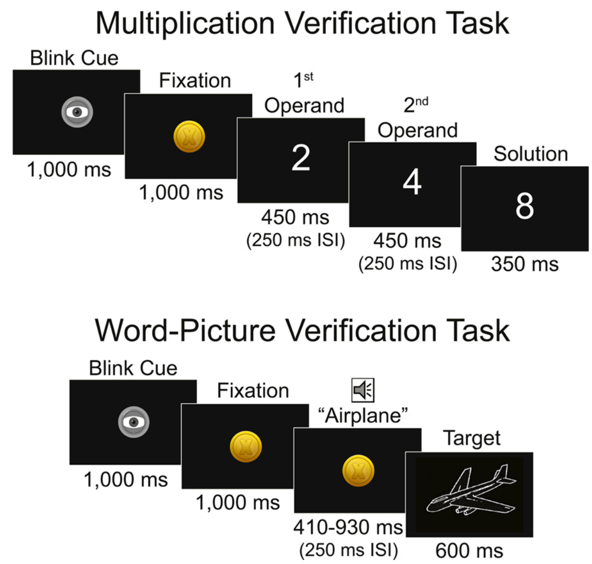

Fig. 1. Trial structure for the multiplication task (top) and for the word-picture verification task (bottom), from left to right across time.

drawing based on the International Picture-Naming Project (Bates et al., 2000; Szekely et al., 2004) (mean number of syllables 2, range 1–4; mean log frequency 1.43; range 0.36–2.75; Davis, 2005). The nouns were prerecorded in a female voice with natural intonation. The sound files were normalized in amplitude and cropped to the onset/offset of each word. Mismatching (semantically unrelated) trials were created by swapping the matching noun with the noun from another trial that matched on animacy (e.g., airplane and umbrella). All pictures and words appeared in both matched and mismatched conditions across participants, so that each participant saw each picture and each word only once. A technical error resulted in 39 match and 41 mismatch trials, for a total of 80 trials per list.

Participants were told a cover story that they were helping aliens learn English by verifying if the spoken noun was the correct name for a given picture or not. The trial structure is shown in Fig. 1. The response period was the same as above. The experiment was divided into 4 blocks of 20 trials. Children earned coins for correct responses, as above.

2.4. Electroencephalogram (EEG) recording and processing

Participants sat alone in a sound attenuating, electrically shielded chamber. A closed-circuit wall-mounted camera was used to monitor the participants, and an intercom was used to communicate while the chamber door was closed. Continuous electroencephalograms (EEG) were recorded from 26 Ag-AgCl sintered active scalp electrodes arranged in a geodesic array in an elastic cap (Electro-Cap International Inc.). EEG signals were amplified using a fully DC-coupled BioSemi ActiveTwo bioamplifier (BioSemi B.V., Amsterdam, Netherlands). Electrodes placed on the outer canthi and under each eye captured horizontal eye movements and blinks, respectively. Electrodes placed over the left and right mastoid processes were used as an off-line reference (average mastoids). All electrode offsets were kept under 25 millivolts. The data were sampled at 256 Hz (2048 Hz with a decimation factor of 1/8) with a fixed first order analog antialiasing filter (-3 dB at 3.6 kHz) (see https://www.biosemi.com).

A combination of EEGLab (version 14.1.2) and ERPLab (version 7.0.0) was used to process the data. Event-related potentials (ERPs) were extracted from the continuous EEG data by time-locking to the onset of the solution or picture and averaged by condition of interest (correct/match or incorrect/mismatch trials) relative to a 100 ms prestimulus baseline. Epochs of raw EEG data (-100 to 900 ms around the onset of the solutions) were inspected visually for eye blinks, movements, and drifts to determine individualized thresholds. Automatic artifact rejection algorithms were applied to exclude epochs from analysis and allow for higher participant retention. Participants were included in analysis if they had a minimum of 15 trials per critical condition (children average: 31 trials, range: 15–55; adults average: 47 trials, range: 20–56). Two 2nd order Butterworth digital filters, high pass (low cutoff) at 0.1 Hz and low pass (high cutoff) at 30 Hz, were applied to the data prior to analysis. All statistical analyses were performed in R (R Core Team, 2013) analyses of variance (ANOVA) were run using the *ezanova* package (Lawrence, 2016). Only trials with accurate responses were included in analyses; trials that ended prior to registering a response were excluded.

3. Results

3.1. Experiment 1a: children

3.1.1. Behavior

Fig. 2 shows the behavioral results for the multiplication and the word-picture tasks. Table 2 summarizes the behavioral results for the multiplication task. Separate ANOVAs were performed with two levels of Correctness (correct, incorrect) for the multiplication task or two levels of Match (match, mismatch) for the word-picture task. RTs were measured from the onset of the solution or picture for trials with accurate responses.

For the multiplication task, mean accuracy was 83 % (range: 50–100 %) and mean RT was 1380 ms (range: 553−2692 ms). There was an effect of Correctness for both accuracy ($F(1,98) = 15.75$, $p < 0.001$) and RT ($F(1,98) = 106.48$, $p < 0.001$), with faster and more accurate responses for correct (accuracy: 85 %; RT: 1306 ms) than incorrect problems (accuracy: 81 %; RT: 1489 ms). For the word-picture task, mean accuracy was 93 % (range: 58–100 %) and mean RT was 1039 ms (range: 603–1549 ms); there was no effect of correctness for either accuracy or RT, likely a ceiling effect. For the children who performed both tasks (n=62), they were significantly faster and more accurate on the word-picture verification task than on the multiplication verification task (RT: $F(1,61) = 22.25$, $p < 0.001$; Accuracy: $F(1,61) = 15.13$, $p < 0.001$).

Between-subject factors of real-world significance were also analyzed. Separate ANOVAs for Grade (3rd, 4th and 5th), Gender (female, male), and Language (monolinguals, bilinguals) were performed for accuracy and RT. There was an effect of Grade on RT for the multiplication task ($F(2,96) = 4.99$, $p < 0.05$), where children were faster in responding to solutions with increasing grade (and increasing age Pearson's $r = 1.00$, $p < 0.01$). No other contrasts were significant for either task or dependent measure.

Pearson's correlations revealed that higher WIAT III raw scores (Wechsler, 2009) and higher working memory scores (numbers reversed, Woodcock et al., 2001) were associated with higher accuracy (WIAT $r = 0.64$, $p < 0.01$; WJ III $r = 0.27$, $p < 0.01$) and faster RTs (WIAT $r = -0.54$, $p < 0.01$; WJ III $r = -0.35$, $p < 0.01$) on the multiplication verification task, validating that the online judgment task was sensitive to arithmetic fluency.

3.1.2. ERPs Waveform morphology

Fig. 5 shows the grand-average ERPs from the vertex electrode, time-locked to the onset of the solution (multiplication task) or drawing (word-picture verification task). For both tasks, sensory components (P1-N1-P2) were followed by a negative-going deflection around 400 ms post-stimulus onset with larger amplitude for incorrect/mismatch than correct/match trials – an N400 effect.

3.1.3. Peak Latency: N400

Only the children who completed both tasks (n = 62) were included in the analyses that directly compared the multiplication and word-picture verification tasks. Peak latency of the N400 effect (incorrect – correct solutions) was identified using mean amplitude of all 26 electrodes between 200 and 600 ms. The N400 effect peaked at 393 ms (SE = 7) on average for the multiplication task and 387 ms (SE = 7) on average for the word-picture verification task. An ANOVA with two levels of Task (multiplication, word-picture) and 26 levels of Electrode revealed no significant main effect of Task ($F(1,61) = 0.57$, $p = 0.45$), indicating that the peak latency of the effect was not significantly different across tasks. In turn, a 200 ms measurement window centered at 390 ms (290−490 ms post stimulus onset) was used for analysis in both tasks.

3.1.4. Mean Amplitude: N400

For the multiplication task, an omnibus ANOVA with two levels of Correctness (correct, incorrect) and 26 levels of Electrode revealed a main effect of Correctness ($F(1,98) = 38.3$, $p < 0.001$) with larger negative amplitude for incorrect (1.64 μV, SE = 0.43) than correct solutions (4.24 μV, SE = 4.24). A significant interaction between Correctness and Electrode ($F(25,74) = 3.14$, $p < 0.001$) was explored with a subsequent distributional analysis, with factors of Correctness (correct, incorrect), Hemisphere (left, right), Laterality (lateral, medial), and Anteriority (prefrontal, frontal, parietal, occipital). This analysis revealed a maximum N400 correctness effect over medio-central electrodes (Fig. 3; Correctness by Anteriority, $F(3,294) = 5.31$, $p < 0.05$; Correctness by Laterality, $F(1,98) = 16.10$, $p < 0.05$). This distribution is consistent with a typical N400 effect (see (Kutas and Federmeier, 2011).

For the word-picture task, an omnibus ANOVA with two levels of Match (match, mismatch) and 26 levels of Electrode revealed a main effect of Match ($F(1,61) = 30.15$, $p < 0.001$), with larger negative amplitude for mismatch (-2.02 μV, SE = 0.84) than match (2.93 μV, SE = 0.87) pictures. A subsequent distributional analysis with the same factors as above was performed to explain an interaction between Match and Electrode ($F(25,28) = 4.67$, $p < 0.001$). Significant interactions of Correctness with Anteriority ($F(3,183) = 4.20$, $p < 0.05$) and Laterality ($F(1,61) = 36.11$, $p < 0.05$) reflected a maximum N400 correctness effect over medio-central electrodes (see Fig. 4 for the topographic plots).

When comparing the effects across tasks (incorrect – correct for the multiplication; mismatch – match for the word-picture task), a distributional analysis with factors of Task (Multiplication, Word-picture), Hemisphere (left, right), Laterality (lateral, medial), and Anteriority (prefrontal, frontal, parietal, occipital) found no main effect of Task ($F(1,61) = 3.17$, $p = 0.08$). A significant interaction of Task with Anteriority ($F(3,59) = 5.77$, $p < 0.01$) reflected a more anterior effect for pictures in the word-picture verification task than digit solutions in the multiplication task (comparison of prefrontal and frontal levels, separately, across task, p values <0.01). This is consistent with studies showing a more anterior distribution of the N400 for pictures than words (Federmeier and Kutas, 2001; Ganis et al., 1996; Mudrik et al., 2010). There was no significant difference in distribution between the effects over the posterior channels (comparison of parietal and occipital, separately, levels across task, p values >0.1), or for any other effects.

3.1.5. Effect of grade

Single factor ANOVAs with Grade (3, 4, 5) as a between-subject factor were conducted to determine the effect of grade on each ERP component. Grade did not significantly modulate mean amplitude of the P1, N1 or N400 for either task ($p > 0.05$). The mean amplitude of the P2 decreased as grade increased on both tasks (multiplication task: $F(2,96) = 5.52$, $p < 0.01$, word-picture task: $F(2,59) = 3.40$, $p < 0.05$), consistent with expected decreases in sensory component amplitude with increasing age (age by P2 amplitude $r = -0.26$, $p < 0.01$; Friedman, 2012); Fig. 6 shows a representative electrode. None of the standardized

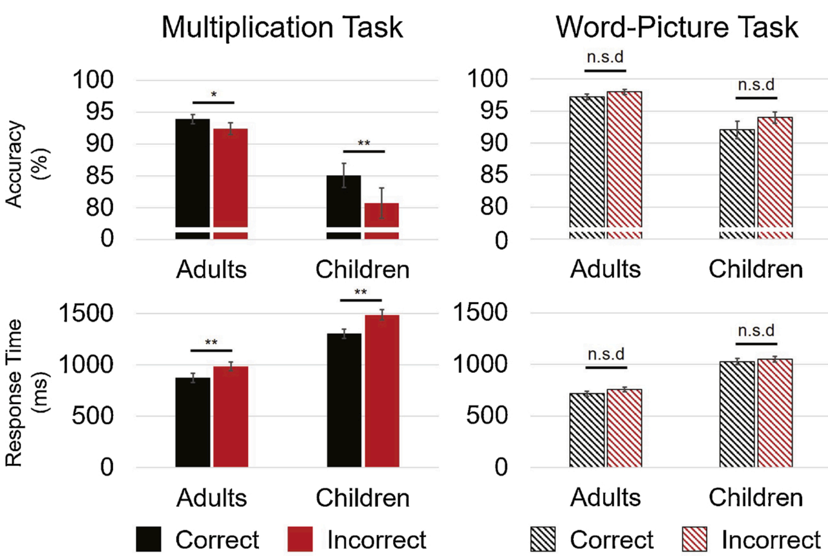

Fig. 2. Behavioral summary for the multiplication verification task (left) and the word-picture verification task (right). Percent accuracy and response time in milliseconds were measured from the onset of the solution/picture. Blacks bars indicate trials ending with correct solutions/matching pictures and red bars indicate trials ending with incorrect solutions/mismatching pictures. *p < 0.05, **p < 0.001. (For interpretation of the colors of this figure, the reader is referred to the web version of this article.)

Table 2
Online response time and accuracy for the multiplication verification task and Pearson's correlation values for each with standardized test scores.

	Children (n = 99)	Adults (n = 60)
Response Time	1379.75 ms (SD = 442.68)	925.10 ms (SD = 331.78)
Math fluency (Multiplication)	r = -0.54**	r = -0.61**
Working memory	r = -0.35**	r = -0.13
Phonological awareness	r = -0.04	r = -0.06
Accuracy	82.90 % (SD = 12.97)	93.15 % (SD = 7.16)
Math fluency (Multiplication)	r = 0.64**	r = 0.64**
Working memory	r = 0.27**	r = 0.29*
Phonological awareness	r = 0.07	r = 0.09

* Significant at p < 0.05.
** Significant at p < 0.01.

measures or task response measures (RT and accuracy) predicted the N400 Correctness effect for either task.[1] However, additional analyses revealed differences in the timing of the onset and offset of the N400 effect across grades in the multiplication task. The N400 effect became significant earlier and ended later for fifth graders than 3rd and 4th graders (effect duration 3rd: 314–414 ms; 4th: 306–440 ms; 5th: 268–490 ms). Mean amplitude and area analyses were run again using the wider 5th grade time window (260−500 ms) to ensure the entire effect was captured. The results did not change: grade did not modulate the N400 effect for mean amplitude ($F(2,96) = 1.32$, $p = 0.27$) or area ($F(2,96) = 0.77$, $p = 0.46$).

3.2. Experiment 1b: adults

Behavior and ERPs were analyzed as above.

3.2.1. Behavior

Fig. 2 summarizes the behavioral results for the multiplication and the word-picture verification tasks. Mean accuracy for the multiplication task was 93 % (range: 66–100 %) and mean RT was 905 ms (range: 481−1939 ms). The correctness effect reached significance for both accuracy ($F(1,59) = 4.22$, $p < 0.05$) and RT ($F(1,59) = 49.57$, $p < 0.001$). Adults were faster and more accurate on correct (accuracy: 94 %; RT: 876 ms) than incorrect problems (accuracy: 92 %; RT: 987 ms). Children were slower and less accurate overall ($F(1,157) = 33.45$, $p < 0.001$ for accuracy; $F(1,157) = 47.07$, $p < 0.001$ for RT), and showed a larger correctness effect for RT ($F(1,157) = 7.81$, $p < 0.01$) but not accuracy compared to adults.

For the word-picture verification task, mean accuracy was 98 % (range: 84–100 %) and mean RT was 736 ms (range: 459–1071 ms). There was no correctness effect for either measure, likely reflecting a performance ceiling. Adults were faster and more accurate overall compared to children ($F(1,120) = 18.01$, $p < 0.001$ for accuracy; $F(1,120) = 76.78$, $p < 0.001$ for RT); there were no differences across groups in the correctness effect for either measure.

Adults were significantly faster and more accurate on the word-picture verification task than on the multiplication verification task (RT: $F(1,59) = 14.01$, $p < 0.001$; Accuracy: $F(1,59) = 32.15$, $p < 0.001$). Pearson's correlations revealed that higher WIAT III raw scores (Wechsler, 2009) were associated with higher accuracy ($r = 0.64$, $p < 0.01$) and faster RTs ($r = -0.61$, $p < 0.01$) on the multiplication verification task, again validating that the online task was sensitive to arithmetic fluency. Higher working memory scores (numbers reversed, Woodcock et al., 2001) were associated with higher accuracy ($r = 0.29$, $p < 0.05$). There were no differences in the adults as a function of gender or language-group for either dependent measure on either task.

3.2.2. ERP waveform morphology

Visual inspection of the grand-average ERPs (Fig. 5) revealed the typical sensory components (P1-N1-P2) in both tasks after the onset of either the solution or the drawing. In the digit task, a subsequent P300 response is visible for correct solutions consistent with Dickson et al. (2018); Dickson and Federmeier (2017); Dickson and Wicha (2019), and Jasinski and Coch (2012). In contrast, in the word-picture task an N400 is visible for both match and mismatch trials peaking around 400 ms after picture onset. For illustration purposes to appropriately reflect the underlying grand average ERPs components visible in the waveforms, the difference waves used for the topographic isovoltage scalp maps highlight the P300 in the digit task by subtracting the grand average ERPs for correct minus incorrect solutions and the N400 in the word-picture task by using the opposite subtraction (mismatch – match).

3.2.3. Peak latency: P300/N400

To determine the onset of the effects of interest in both tasks, the

[1] Linear Mixed Effects Regression (LMER) models were used to examine if individual variability in the ERP data could be explained by grade. Results were similar to the ANOVA analysis, with grade not predicting the N400 effect for either task.

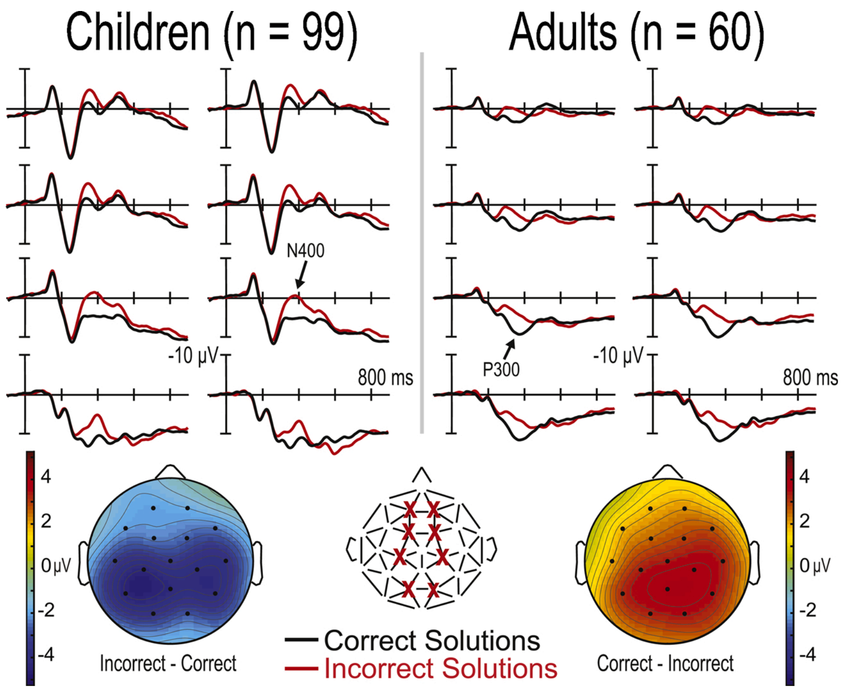

Fig. 3. Grand average ERPs from the multiplication verification task for correct (black lines) and incorrect (red lines) solutions in children (left) and adults (right). X-axis is time in milliseconds with 0 marking the onset of the solution and the Y-axis is voltage in microvolts with negative plotted up. The electrodes plotted are represented with "X" on the head plot. Topographic plots represent the effect between 290-490 ms in children and 255-455 ms in adults. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

average peak latency of the effect (incorrect – correct solutions) was measured between 200 and 600 ms. The P300 effect peaked at 352 ms (SE = 5) for the multiplication task and the N400 effect peaked at 388 ms (SE = 6) for the word-picture task. An ANOVA with two levels of Task (multiplication, word-picture) and 26 levels of Electrode showed a main effect of Task ($F(1,59) = 25.37$, $p < 0.001$). Namely, the P300 effect in the multiplication task peaked significantly earlier than the N400 effect in the word-picture task, consistent with the typical timing of these effects. Accordingly, the measurement windows for the analysis of mean amplitude were centered around these peak values: 255−455 ms for the multiplication task and 290−490 ms for the word-picture task.

3.2.4. Mean amplitude: P300/N400

For the multiplication task, an omnibus ANOVA with two levels of Correctness (correct, incorrect) and 26 levels of Electrode revealed a main effect of Correctness ($F(1,59) = 89.25$, $p < 0.001$), with more positive amplitude for correct (5.08 µV, SE = 0.36) than incorrect solutions (2.7 µV, SE = 0.36). An interaction between Correctness and Electrode ($F(25,35) = 4.61$, $p < 0.001$) was inspected with a subsequent distributional analysis, using the same factors as in Experiment 1a (see section 3.1.4). Significant interactions of Correctness with Anteriority ($F(3,177) = 15.54$, $p < 0.05$), Laterality ($F(1,59) = 60.22$, $p < 0.05$) and Hemisphere ($F(1,59) = 15.08$, $p < 0.05$) reflected a maximum effect over

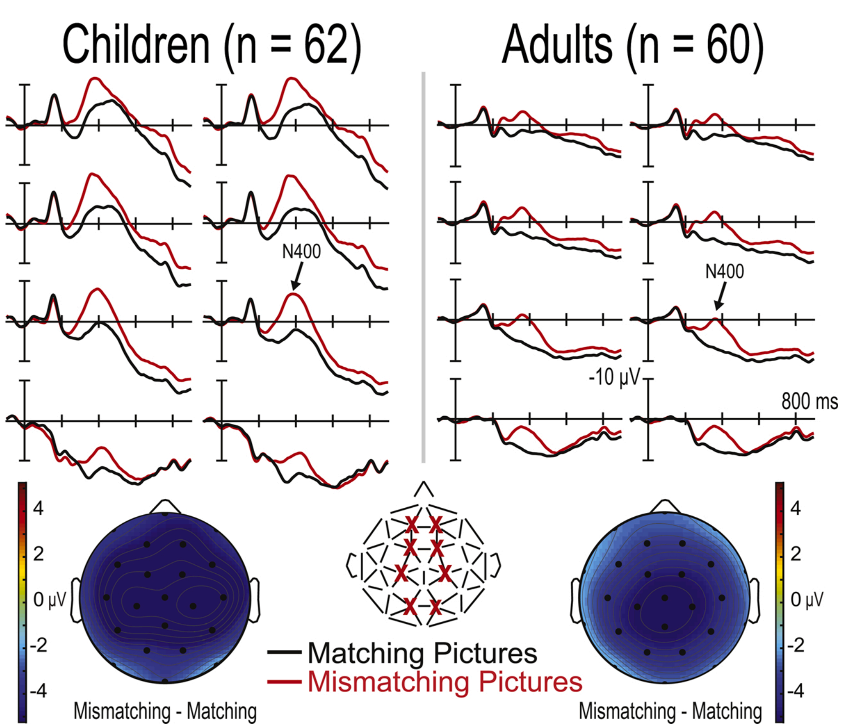

Fig. 4. Grand average ERPs from the word-picture verification task for matching (black lines) and mismatching (red lines) pictures in children (left) and adults (right). X-axis is time in milliseconds with 0 marking the onset of the picture and the Y-axis is voltage in microvolts with negative plotted up. The electrodes plotted are represented with "X" on the head plot. Topographic isovoltage scalp maps represent the effect between 290-490 ms in both children and adults. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

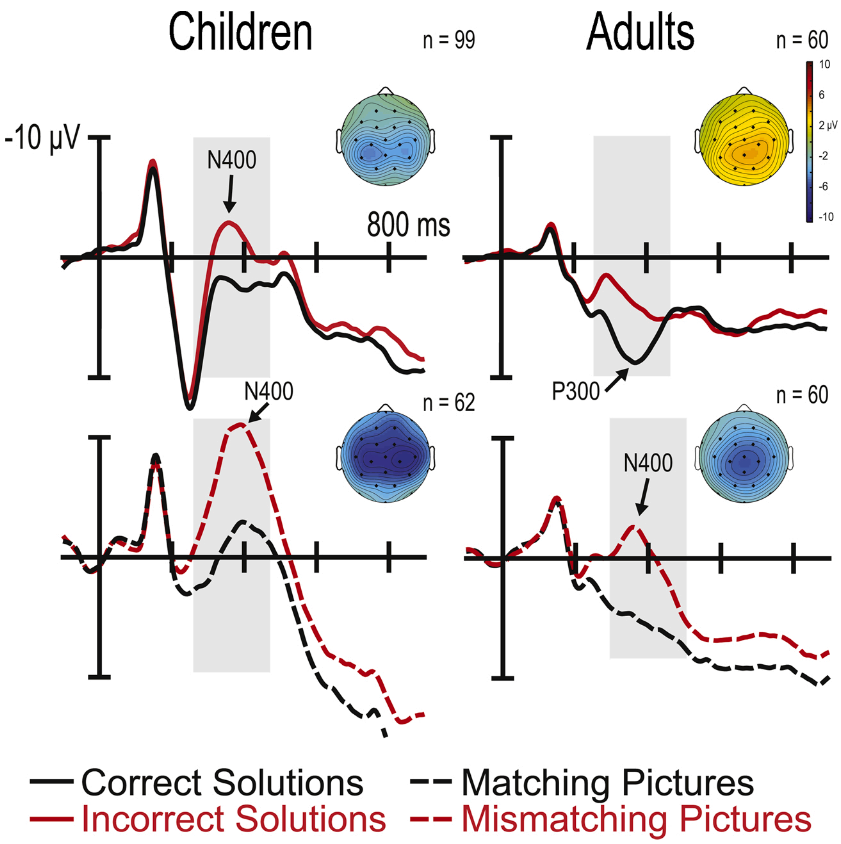

Fig. 5. Summary of the effects at the vertex electrode (MiCe) showing the grand average ERPs for the multiplication task (top) and the word-picture verification task (bottom) in both children (left) and adults (right). Black lines represent the brain response to correct solutions and red lines to incorrect solutions. Dotted black lines represent the brain response to matching trials and red dotted lines to mismatching trials. Grey shaded areas represent the window of analysis (290 to 490 ms for children in both tasks and for adults in the word-picture task; 255 to 455 ms for adults in the multiplication task). Topographic maps of the effects observed in these time windows are represented next to each ERP trace. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

medial-occipital electrodes, typical of the more posterior distribution of the target P300 (Polich, 2007, 2012) (Fig. 3).

For the word-picture task, an omnibus ANOVA with two levels of Match (match, mismatch) and 26 levels of Electrode revealed a main effect of Match ($F(1,59) = 118.45$, $p < 0.05$), with more negative amplitude for mismatched (0.79 µV, SE = 0.50) than matched (4.21 µV, SE = 0.57) pictures. An interaction between Match and Electrode ($F(25,35) = 8.17$, $p < 0.001$) was explored with a distributional analysis, as before. A significant interaction of Correctness with Laterality ($F(1,59) = 184.83$, $p < 0.05$) was driven by a larger effect at medial than lateral sites, consistent with an N400 congruency effect (Kutas and Federmeier, 2011) (Fig. 4).

When comparing the effects across tasks (correct – incorrect for the multiplication; mismatch – match for the word-picture task), a distributional analysis with factors of Task (Multiplication, Word-picture), Hemisphere (left, right), Laterality (lateral, medial), and Anteriority (prefrontal, frontal, parietal, occipital) found a main effect of Task ($F(1,59) = 189.00$, $p < 0.01$), revealing that the amplitude of the N400 effect for the word-picture task was significantly larger than the effect in the multiplication task. There was also an interaction of Task with Anteriority ($F(2,59) = 5.77$, $p < 0.01$) such that the effect in multiplication task was larger in amplitude at the posterior channels (parietal and occipital levels) compared to the anterior channels (prefrontal and frontal levels), which is in line with a typical target P300 distribution (Polich, 2007, 2012). The word-picture task effect was larger across all levels of anteriority compared to the multiplication task effect (p values <0.01 across all levels). No other effects reached significance.

4. Discussion

This study aimed to determine if children and adults engage similar neurocognitive processes when verifying simple multiplication facts. Both adults and children were faster and more accurate when judging correct than incorrect solutions. Overall, adults were faster and more accurate than children (Table 2). Critically, although children had high accuracy in their performance on the math task (~83 %), they did not show the adult-like brain pattern. The ERP results for trials with accurate responses (i.e., participant made the correct judgement) revealed that successful performance on the task was supported by distinct neurocognitive processes across the two populations. Therefore, a developmental shift in arithmetic processing from childhood to adulthood has indeed been overlooked in the literature.

In children, both correct and incorrect solutions elicited an N400, with larger amplitude for contextually unsupported (incorrect) solutions. The word-picture verification task also generated an N400 effect, with larger amplitudes relative to arithmetic facts given the richer semantic content elicited by words/pictures, especially over frontal electrodes which is typical for pictures (Federmeier and Kutas, 2001; Ganis et al., 1996; Mudrik et al., 2010). Critically, the timing of the N400 effects did not differ across tasks (390 ms for both), further supporting that both tasks elicited comparable N400 responses. Therefore, children appear to prepare meaning-level expectations in both tasks and respond to the solutions according to the semantic congruency of their continuation for each sequence of numbers (Jost et al., 2003; Niedeggen et al., 1999; Prieto-Corona et al., 2010). Like children, adults also exhibited an N400 effect in the word-picture task, with reduced N400 amplitude for contextually supported (match) trials. This N400 effect was smaller in amplitude compared to children, which is consistent with a developmental shift toward smaller N400 effects with increasing age (Holcomb et al., 1992; Kutas and Iragui, 1998). For multiplication problems, however, correct solutions elicited a positive-going ERP in adults, not an N400.

The morphology and timing of the adult response to correct solutions is consistent with a typical target P300. The P300 is composed of subcomponents that are sensitive to different cognitive demands. The "oddball" P300 is typically a modulation of the P3b, and reflects categorical decision making with larger amplitude for easier decisions. In contrast, the P3a, also called the novelty P300, is modulated by attention and distractors (see Polich, 2007, 2012). Here, the modulation on the P300 amplitude for correct solutions is likely driven by the categorical decision of correctness (P3b), with the correct solutions being easier to categorize. This finding implies that adults processed the correct solutions as targets for overlearned and rapidly categorized problems and did not rely on meaning-level representations as the children did (i.e., no reduced N400 for correct relative to incorrect solutions). It is worth noting that Prieto-Corona et al. (2010), who argued for an N400 effect in both children and adults, showed similar ERP morphology and timing differences across groups as in the current study. Indeed, the latency measurements for their negativity peaked at ~400 ms for children, consistent with an N400, and ~290 ms for adults, consistent with a P300. Critically, this finding is also consistent with the more recent ERP studies that have reinterpreted the arithmetic correctness effects in adults from an N400 effect to a P300 effect (Dickson et al., 2018; Dickson and Federmeier, 2017; Jasinski and Coch, 2012).

To be clear, arguing for a discrete dichotomy between children and adults, with only children eliciting N400 s and only adults eliciting P300 s, would be misguided. Given that any potentially meaningful stimulus can elicit an N400 (Kutas and Hillyard, 1980), including numbers represented as digits (Dickson and Federmeier, 2018), it is probable that the adults generated small N400 s to every digit as they appeared. Indeed, under the right circumstances, namely when the problems can be treated like language, adults can elicit robust N400 effects to multiplication problems (Dickson et al., 2018; Martinez-Lincoln et al., 2015; Salillas and Wicha, 2012). However, the contribution of the N400 generated by each digit on the ERP waveform morphology, and on the correctness effect itself, appears to be negligible in this and other studies (see Dickson and Federmeier, 2017; Jasinski and Coch, 2012). Similarly, both the word-picture verification and the multiplication verification tasks are categorical and require decision-making. Therefore, it is probable that children generate decision-related P300 s in this task like in other studies with children (Moore et al., 2014; Riggins and Scott, 2020). The critical finding, however, is that the arithmetic correctness effect is driven predominantly by a modulation on the N400 for children and by a modulation of the P300 for adults, alluding to a developmental transition or cognitive tradeoff. Moreover, we propose that this transition is gradual and fluid, with the possibility of reverting back when task demands change (e.g., Dickson et al., 2018).

For this large sample of children, we also analyzed standardized metrics as predictors for performance and the brain measures. Offline measures of math fluency (i.e., from the WIAT) correlated with both accuracy and response time on the multiplication task in children, confirming the validity of the multiplication task. Children responded faster on the multiplication task as grade level (and age) increased but did not improve in accuracy. Given this improvement in response speed, it is reasonable to ask if an experience-based change was also measurable in the ERPs. However, the N400 does not typically correlate with speed of responding, which is the only real time behavioral measure that improved with grade (Heinze et al., 1998). So, it is perhaps not surprising that there was no significant change in the mean amplitude of the N400 effect with grade (or age)[2] . Similarly, the waveform morphology of the ERP response to correct solutions in children did not suggest a shift from an N400 to a P300 as grade level increased (Fig. 6), although perhaps the increase in onset and offset of the N400 effect for 5th graders may allude to the beginning of an underlying morphological change. These findings suggest that, unlike adults, children do not quickly or confidently recognize correct solutions as target items, and instead process arithmetic facts for meaning. This previously unreported ERP difference between children and adults appears to resolve at some point after fifth grade. Future research will need to determine at what point in development, and under what circumstances, children begin to show the adult-like brain response to simple arithmetic.

We hypothesize that the transition from the child-like N400 effect to the adult-like P300 effect is driven by experience rather than a maturational change. Maturational changes are clearly observable on earlier components, namely the P2 (Fig. 6), but no significant modulation occurs on the N400 with age or grade. The shift from N400 to P300 may be analogous to the transition observed in adult second language learning, where in early stages of learning adults show an N400 to grammatical violations, then as fluency in the language increases, a native-like positivity is observed to the same sentences (McLaughlin et al., 2010; Tanner et al., 2013). Another possible explanation for our results is that this difference in brain responses between children and adults is generational rather than developmental.

Early math education typically emphasizes rote memorization for learning multiplication tables. However, more recently a greater emphasis has been placed on conceptual understanding, given that memorization alone can lead to challenges in understanding later math concepts, such as division (Dubé and Robinson, 2018; Sullivan and McDuffie, 2009). For example, in 2014 the National Council of Teachers of Mathematics in the U.S.A. called for a change in core curriculum toward more conceptually based learning methods, such as the concrete-representational-abstract (CRA) sequence that uses multiple representations to help children understand math concepts (NCTM, 2014). Although Texas, where this study was conducted, does not follow the national curriculum, Texas public schools did adopt conceptualized learning (Texas Essential Knowledge and Skills for Mathematics: implemented in the 2014–2015 school year). Therefore, it is possible that our sample of children may have learned their facts with more of a conceptual understanding than the young adults in our sample. This could, in turn, have led to a difference in engagement of conceptual (meaning) level information when verifying multiplication facts on our task, potentially leading to an N400 in the younger but not the older populations.

Importantly, however, both children and adults reported primarily using retrieval to solve multiplications in our offline strategy assessment, which was based on research showing that self-assessment is a valid measure of strategy use (Siegler and Stern, 1998). Moreover, our ERP results are consistent with studies conducted with children and adults from varied demographics, including populations outside the USA (Cerda et al., 2019; Dickson et al., 2018; Dickson and Federmeier, 2017; Jasinski and Coch, 2012; Moore et al., 2014; Prieto-Corona et al., 2010). This would suggest that it is not (only) the learning methods of our samples that lead to this difference across populations. Future experiments using longitudinal studies could determine if the N400-P300 difference across our groups is related to the strength of conceptual understanding of the math facts.

Also emerging from our results is the question of what this N400 effect reflects in children, more broadly. The extensive literature on the N400 suggests that it is an index of the state of the memory system when processing any meaningful or potentially meaningful stimulus (see Kutas and Federmeier, 2011 for a review). Although its neural source may vary based on the type of information being processed, the N400 is likely generated by temporal lobe sources in the brain (Lau and Namyst, 2019; Lau et al., 2008). However, researchers have argued that the meaning of numbers, in the sense of numerical quantity, is dependent on the parietal lobe (Dehaene and Cohen, 1995; Dehaene et al., 2003). For example, both the triple code model (TCM; Dehaene, 1992) and the encoding complex model (ECM; Campbell and Clark, 1988), which have been based on adult data, propose a magnitude representation that is separate from the representation of arithmetic facts and number words. Use of

[2] Although P300 latency has been shown to correlate with response latency on tasks that involve simple decisions (Kutas et al., 1977; Polich, 2012), this relationship becomes less reliable with more complex decisions (Folstein and Van Petten, 2011). Similarly, there was no correlation between response time and P300 latency in the adult multiplication task (all p values >0.10).

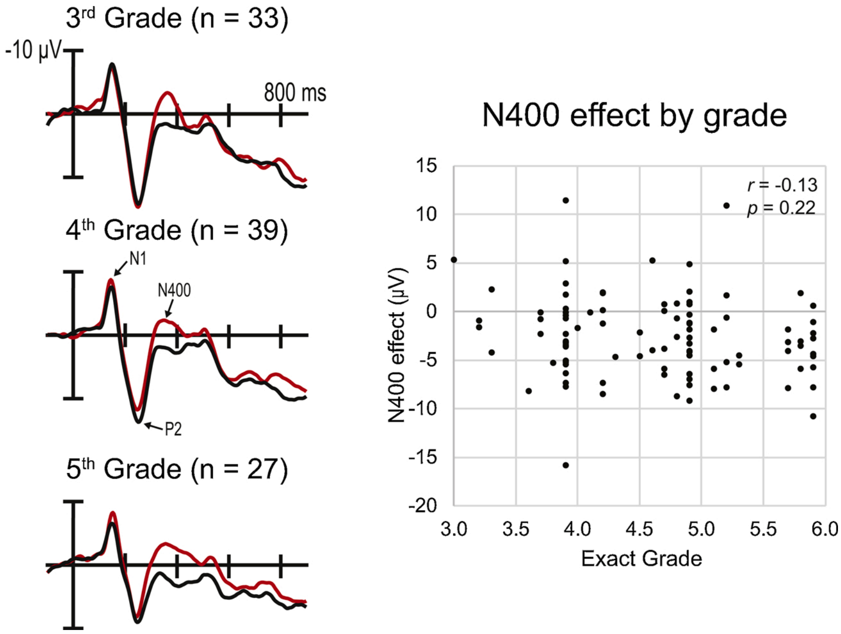

Fig. 6. Grand average ERPs (left) shown at the vertex electrode (MiCe) for the multiplication verification task broken down by grade level (top to bottom). Black lines represent the brain response to correct solutions and red lines to incorrect solutions. The scatter plot (right) shows the N400 correctness effect (mean amplitude for incorrect minus correct solutions; y-axis) by grade (x-axis). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

this so-called magnitude code arguably relies on the parietal cortex, which is engaged during estimation or calculation (Ansari, 2008; Cohen Kadosh et al., 2008; Dehaene et al., 2003). In contrast, access to memorized arithmetic facts is thought to be mediated through verbal memories of idiomized representations of the facts (see Dehaene and Cohen, 1995, 1997). Stroke patient data suggest that these systems may be dissociable, with temporal lobe damage leading selectively to retrieval problems and parietal lobe damage selectively leading to estimation or computation problems (Cohen et al., 2000; Delazer and Benke, 1997; Lemer et al., 2003). And, in children, a developmental shift for multiplication facts from relying more on parietal cortex to relying more on temporal cortex is thought to parallel the transition from calculating (meaning) to retrieving the facts (Prado et al., 2014).

The conclusions drawn from these studies, which use measures of brain activity with poor temporal resolution, may conflict with the present ERPs findings, which reflect real time brain activity. However, because the N400 itself reflects meaning-level processing more broadly, it cannot distinguish between different sources of meaning (e.g., numerosity versus verbal memory). Therefore, the N400 effect in children could potentially implicate the parietal lobe as a novel N400 source, if conceptual processing of the numerical operation is responsible for the effect, or alternatively, implicate the temporal lobe in meaningful processing of arithmetic fact retrieval, contrary to what is suggested by the models of arithmetic. Furthermore, it is notable that adults, who are thought to rely heavily on verbal memory, do not seem to show N400 effects in this study, but can show N400 effects when the task encourages language-like processing of the math facts (Dickson et al., 2018). Therefore, the current findings in children raise important theoretical and empirical questions. Is this arithmetic N400 effect, which indexes meaning-level processes, generated in parietal cortex? Or, do children access meaning for arithmetic facts from temporal cortex, similar to words? If the latter, what then happened to that meaning in the adult? Answers to these questions in children would inform current models of mathematical cognition, which have been based primarily on findings from adults.

5. Conclusion

Arithmetic retrieval fluency is associated with better performance in advanced math (Geary et al., 2013; Price et al., 2013; Siegler et al., 2012). This study measured the neurocognitive processes supporting multiplication fact verification in children and adults, and compared the results to a language task (word-picture verification). Children showed an N400 effect in both tasks, with larger amplitude for the incorrect than correct solutions, and for mismatching than matching pictures. Adults also showed a reduced N400 response for matching relative to non-matching pictures in the word-picture verification task, but instead showed a qualitatively different ERP in the multiplication task, namely a P300 to correct solutions. These results suggest that the faster response times and better accuracy in adults is not simply an increase in efficiency compared to children, but rather the engagement of a different cognitive process. While children rely primarily on meaning-level processes for verifying simple multiplication problems, the P300 in adults suggests that they treat the solutions as potential targets in over-rehearsed problems and categorize them efficiently. This study adds to the current math cognition literature by demonstrating that the transition to an adult-like brain response is more gradual than suggested by previous ERP studies, with a progressive developmental shift happening beyond fifth grade (LeFevre et al., 1996; Siegler, 1996). Moreover, these findings highlight the need for current models of math cognition to account for neurocognitive evidence from children to fully understand the nature of arithmetic in the brain.

Funding

This study was supported by awards R21HD079884 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, and BRAIN EAGER award 1451032 from the National Science Foundation. This project was funded (in-part) by the University of Texas at San Antonio, Office of the Vice President for Research, Economic Development, and Knowledge Enterprise.

Declaration of Competing Interest

The listed authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflict of interest.

Acknowledgements

This work received computational support from the UTSA's HPC Cluster Shamu, operated by the Office of Information Technology. We thank J. Momsen for assistance developing the spoken number words and the language task, and A.S. Niembro, the voice of the nouns. Thanks also to our participant recruitment team, B. Obinyan, A. Godinez, and P. Montufar. We thank all of the subjects and their families for their participation.

References

Ansari, D., 2008. Effects of development and enculturation on number representation in the brain. Nat. Rev. Neurosci. 9 (4) https://doi.org/10.1038/nrn2334.
Bates, E., et al., 2000. Introducing the CRL international picture-naming project (CRL-IPNP). CRL Newsletter 12 (1), 4–14.
Campbell, J.I.D., Clark, J.M., 1988. An Encoding-Complex View of Cognitive Number Processing: Comment on McCloskey, Sokol, and Goodman (1986). J. Exp. Psychol. Gen. https://doi.org/10.1037/0096-3445.117.2.204.
Campbell, J.I.D., Graham, D.J., 1985. Mental multiplication skill: structure, process, and acquisition. Can. J. Psychol. Can. Psychol. 39 (2), 338–366. https://doi.org/10.1037/h0080065.
Cerda, V.R., Grenier, A.E., Wicha, N.Y.Y., 2019. Bilingual children access multiplication facts from semantic memory equivalently across languages: evidence from the N400. Brain Lang. https://doi.org/10.1016/j.bandl.2019.104679.
Cipolotti, L., Butterworth, B., Denes, G., 1991. A specific deficit for numbers in a case of dense acalculia. Brain 114 (6). https://doi.org/10.1093/brain/114.6.2619.
Cohen, L., Dehaene, S., Chochon, F., Lehéricy, S., Naccache, L., 2000. Language and calculation within the parietal lobe: a combined cognitive, anatomical and fMRI study. Neuropsychologia. https://doi.org/10.1016/S0028-3932(00)00038-5.
Cohen Kadosh, R., Lammertyn, J., Izard, V., 2008. Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. Prog. Neurobiol. 84 (2) https://doi.org/10.1016/j.pneurobio.2007.11.001.
Davis, C.J., 2005. N-watch: a program for deriving neighborhood size and other psycholinguistic statistics. Behav. Res. Methods 37 (1), 65–70. Retrieved from. http://www.ncbi.nlm.nih.gov/pubmed/16097345.
Dehaene, S., 1992. Varieties of numerical abilities. Cognition. https://doi.org/10.1016/0010-0277(92)90049-N.
Dehaene, S., Cohen, L., 1995. Towards an anatomical and functional model of number processing. Math. Cogn. 1 (1), 83–120.
Dehaene, S., Cohen, 1997. Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. Cortex 33 (2). https://doi.org/10.1016/S0010-9452(08)70002-9.
Dehaene, S., Piazza, M., Pinel, P., Cohen, L., 2003. Three parietal circuits for number processing. Cogn. Neuropsychol. https://doi.org/10.1080/02643290244000239.
Delazer, M., Benke, T., 1997. Arithmetic facts without meaning. Cortex 20 (3–6). https://doi.org/10.1016/S0010-9452(08)70727-5.
Dickson, D.S., Federmeier, K.D., 2017. The language of arithmetic across the hemispheres: an event-related potential investigation. Brain Res. 1662, 46–56. https://doi.org/10.1016/j.brainres.2017.02.019.
Dickson, D.S., Federmeier, K.D., 2018. Your favorite number is special (to you): evidence for item-level differences in retrieval of information from numerals. Neuropsychologia 117. https://doi.org/10.1016/j.neuropsychologia.2018.05.018.
Dickson, D.S., Wicha, N.Y.Y., 2019. P300 amplitude and latency reflect arithmetic skill: an ERP study of the problem size effect. Biol. Psychol. 148 https://doi.org/10.1016/j.biopsycho.2019.107745.
Dickson, D.S., Cerda, V.R., Beavers, R.N., Ruiz, A., Castañeda, R., Wicha, N.Y.Y., 2018. When 2 × 4 is meaningful: the N400 and P300 reveal operand format effects in multiplication verification. Psychophysiology 55 (11). https://doi.org/10.1111/psyp.13212.
Dubé, A.K., Robinson, K.M., 2018. Children's understanding of multiplication and division: insights from a pooled analysis of seven studies conducted across 7 years. Br. J. Dev. Psychol. 36 (2) https://doi.org/10.1111/bjdp.12217.
Federmeier, K.D., Kutas, M., 2001. Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing. J. Exp. Psychol. Learn. Mem. Cogn. 27 (1) https://doi.org/10.1037/0278-7393.27.1.202.
Folstein, J.R., Van Petten, C., 2011. After the P3: late executive processes in stimulus categorization. Psychophysiology 48 (6). https://doi.org/10.1111/j.1469-8986.2010.01146.x.
Friedman, D., 2012. The components of aging. The Oxford Handbook of Event-Related Potential Components. https://doi.org/10.1093/oxfordhb/9780195374148.013.0243.
Friedrich, M., Friederici, A.D., 2004. N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. J. Cogn. Neurosci. 16 (8) https://doi.org/10.1162/0898929042304705.
Ganis, G., Kutas, M., Sereno, M.I., 1996. The search for "Common sense": an electrophysiological study of the comprehension of words and pictures in reading. J. Cogn. Neurosci. 8 (2) https://doi.org/10.1162/jocn.1996.8.2.89.
Geary, D.C., 2011. Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. Dev. Psychol. 47 (6) https://doi.org/10.1037/a0025510.
Geary, D.C., Hoard, M.K., Nugent, L., Bailey, D.H., 2013. Adolescents' functional numeracy is predicted by their school entry number system knowledge. PLoS One 8 (1). https://doi.org/10.1371/journal.pone.0054651.
Heinze, H.J., Muente, T.F., Kutas, M., 1998. Context effects in a category verification task as assessed by event-related brain potential (ERP) measures. Biol. Psychol. 47 (2), 121–135. https://doi.org/10.1016/S0301-0511(97)00024-0.
Holcomb, P.J., Coffey, S.A., Neville, H.J., 1992. Visual and auditory sentence processing: a developmental analysis using event-related brain potentials. Dev. Neuropsychol. 8 (2–3), 203–241. https://doi.org/10.1080/87565649209540525.
Hollingshead, A.B., 2011. Four factor index of social status. Yale Journal of Sociology.
Jasinski, E.C., Coch, D., 2012. ERPs across arithmetic operations in a delayed answer verification task. Psychophysiology 49 (7). https://doi.org/10.1111/j.1469-8986.2012.01378.x.
Jordan, N.C., Hanich, L.B., Kaplan, D., 2003. Arithmetic fact mastery in young children: a longitudinal investigation. J. Exp. Child Psychol. 85 (2) https://doi.org/10.1016/S0022-0965(03)00032-8.
Jost, K., Hennighausen, E., Rösler, F., 2003. Comparing arithmetic and semantic fact retrieval: effects of problem size and sentence constraint on event-related brain potentials. Psychophysiology 40 (6). https://doi.org/10.1111/1469-8986.00119_41_1.
Kawashima, R., Taira, M., Okita, K., Inoue, K., Tajima, N., Yoshida, H., et al., 2004. A functional MRI study of simple arithmetic - A comparison between children and adults. Cogn. Brain Res. 18 (3) https://doi.org/10.1016/j.cogbrainres.2003.10.009.
Kutas, M., Federmeier, K.D., 2011. Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP), vol. 62. SSRN. https://doi.org/10.1146/annurev.psych.093008.131123.
Kutas, M., Hillyard, S.A., 1980. Event-related brain potentials to semantically inappropriate and surprisingly large words. Biol. Psychol. 11 (2), 99–116. https://doi.org/10.1016/0301-0511(80)90046-0.
Kutas, M., Iragui, V., 1998. The N400 in a semantic categorization task across 6 decades. Electroencephalography and Clinical Neurophysiology - Evoked Potentials 108 (5). https://doi.org/10.1016/S0168-5597(98)00023-9.
Kutas, M., Mccarthy, G., Donchin, E., 1977. Augmenting mental chronometry: the p300 as a measure of stimulus evaluation time. Science 197 (4305). https://doi.org/10.1126/science.887923.
Lau, E.F., Namyst, A., 2019. fMRI evidence that left posterior temporal cortex contributes to N400 effects of predictability independent of congruity. Brain Lang. 199, 104697. https://doi.org/10.1016/j.bandl.2019.104697.
Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de) constructing the N400. Nat. Rev. Neurosci. 9 (12), 920–933. https://doi.org/10.1038/nrn2532.
Lawrence, M., 2016. Package "ez." R Topics Documented.
LeFevre, J.A., Sadesky, G.S., Bisanz, J., 1996. Selection of procedures in mental addition: reassessing the problem size effect in adults. J. Exp. Psychol. Learn. Mem. Cogn. 22 (1) https://doi.org/10.1037/0278-7393.22.1.216.
Lemer, C., Dehaene, S., Spelke, E., Cohen, L., 2003. Approximate quantities and exact number words: dissociable systems. Neuropsychologia 41 (14). https://doi.org/10.1016/S0028-3932(03)00123-4.
Marian, V., Blumenfeld, H.K., Kaushanskaya, M., 2007. The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. J. Speech Lang. Hear. Res. 50 (4), 940–967. https://doi.org/10.1044/1092-4388(2007/067).
Martinez-Lincoln, A., Cortinas, C., Wicha, N.Y.Y., 2015. Arithmetic memory networks established in childhood are changed by experience in adulthood. Neurosci. Lett. 584, 325–330. https://doi.org/10.1016/j.neulet.2014.11.010.
McCloskey, M., Macaruso, P., Whetstone, T., 1992. The functional architecture of numerical processing mechanisms: defending the modular model. Adv. Psychol. 91, 493–537. https://doi.org/10.1016/S0166-4115(08)60895-X.
McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., Osterhout, L., 2010. Brain potentials reveal discrete stages of L2 grammatical learning. Lang. Learn. 60 https://doi.org/10.1111/j.1467-9922.2010.00604.x.
Moore, R.D., Drollette, E.S., Scudder, M.R., Bharij, A., Hillman, C.H., 2014. The influence of cardiorespiratory fitness on strategic, behavioral, and electrophysiological indices of arithmetic cognition in preadolescent children. Front. Hum. Neurosci. 8. https://doi.org/10.3389/fnhum.2014.00258.
Mudrik, L., Lamy, D., Deouell, L.Y., 2010. ERP evidence for context congruity effects during simultaneous object-scene processing. Neuropsychologia 48 (2). https://doi.org/10.1016/j.neuropsychologia.2009.10.011.
Naccache, L., Dehaene, S., 2001. The priming method: imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. Cereb. Cortex 11 (10). https://doi.org/10.1093/cercor/11.10.966.
Niedeggen, M., Rösler, F., Jost, K., 1999. Processing of incongruous mental calculation problems: evidence for an arithmetic N400 effect. Psychophysiology 36 (3), 307–324. https://doi.org/10.1017/S0048577299980149.
Núñez-Peña, M., Cortiñas, M., Escera, C., 2006. Problem size effect and processing strategies in mental arithmetic. Neuroreport 17 (4). https://doi.org/10.1097/01.wnr.0000203622.24953.c2.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9 (1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4.
Peters, L., De Smedt, B., 2018. Arithmetic in the developing brain: a review of brain imaging studies. Dev. Cogn. Neurosci. (30) https://doi.org/10.1016/j.dcn.2017.05.002.
Polich, J., 1987. Task difficulty, probability, and inter-stimulus interval as determinants of P300 from auditory stimuli. Electroencephalography and Clinical Neurophysiology/ Evoked Potentials 68 (4). https://doi.org/10.1016/0168-5597(87)90052-9.
Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118 (October), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019.

Polich, J., 2012. Neuropsychology of P300. The Oxford Handbook of Event-Related Potential Components. https://doi.org/10.1093/oxfordhb/9780195374148.013.0089.

Prado, J., Mutreja, R., Booth, J.R., 2014. Developmental dissociation in the neural responses to simple multiplication and subtraction problems. Dev. Sci. 17 (4), 537–552. https://doi.org/10.1111/desc.12140.

Price, G.R., Mazzocco, M.M.M., Ansari, D., 2013. Why mental arithmetic counts: brain activation during single digit arithmetic predicts high school math scores. J. Neurosci. 33 (1) https://doi.org/10.1523/JNEUROSCI.2936-12.2013.

Prieto-Corona, B., Rodríguez-Camacho, M., Silva-Pereyra, J., Marosi, E., Fernández, T., Guerrero, V., 2010. Event-related potentials findings differ between children and adults during arithmetic-fact retrieval. Neurosci. Lett. 468 (3), 220–224. https://doi.org/10.1016/j.neulet.2009.10.094.

Riggins, T., Scott, L.S., 2020. P300 development from infancy to adolescence. Psychophysiology 57 (7). https://doi.org/10.1111/psyp.13346.

Salillas, E., Wicha, N.Y.Y., 2012. Early learning shapes the memory networks for arithmetic: evidence from brain potentials in bilinguals. Psychol. Sci. 23 (7), 745–755. https://doi.org/10.1177/0956797612446347.

Siegler, R.S., 1988. Strategy choice procedures and the development of multiplication skill. J. Exp. Psychol. Gen. 117 (3), 258–275. https://doi.org/10.1037/0096-3445.117.3.258.

Siegler, R.S., 1996. Emerging minds: the process of change in children's thinking. Choice Rev. Online. https://doi.org/10.5860/choice.34-5984.

Siegler, R.S., Stern, E., 1998. Conscious and unconscious strategy discoveries: a microgenetic analysis. J. Exp. Psychol. Gen. 127 (4) https://doi.org/10.1037/0096-3445.127.4.377.

Siegler, R.S., Duncan, G.J., Davis-Kean, P.E., Duckworth, K., Claessens, A., Engel, M., et al., 2012. Early predictors of high school mathematics achievement. Psychol. Sci. 23 (7) https://doi.org/10.1177/0956797612440101.

Snodgrass, J.G., Vanderwart, M., 1980. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J. Exp. Psychol. Hum. Learn. 6 (2) https://doi.org/10.1037/0278-7393.6.2.174.

Sullivan, A., McDuffie, A., 2009. Connecting multiplication to contexts and language In: Teaching Children Mathematics, 15th ed., Vol. 15, pp. 502–512

Sutton, S., Braren, M., Zubin, J., John, E.R., 1965. Evoked-potential correlates of stimulus uncertainty. Science 150 (3700). https://doi.org/10.1126/science.150.3700.1187.

Sutton, S., Ruchkin, D.S., Munson, R., Kietzman, M.L., Hammer, M., 1982. Event-related potentials in a two-interval forced-choice detection task. Percept. Psychophys. 32 (4) https://doi.org/10.3758/BF03206242.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., et al., 2004. A new on-line resource for psycholinguistic studies. J. Mem. Lang. 51 (2) https://doi.org/10.1016/j.jml.2004.03.002.

Tanner, D., McLaughlin, J., Herschensohn, J., Osterhout, L., 2013. Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. Bilingualism 16 (2). https://doi.org/10.1017/S1366728912000302.

Team, R, 2013. R Development Core Team. R: A Language and Environment for Statistical Computing.

The National Council of Teachers of Mathematics, 2014. Principles to Action: Ensuring Mathematical Success for All.

Wechsler, D., 2009. Wechsler Individual Achievment Test, 3rd ed. Psychological Corporation, San Antonio, TX.

Wicha, N.Y.Y., Moreno, E.M., Kutas, M., 2003. Expecting gender: an event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. Cortex 39 (3). https://doi.org/10.1016/S0010-9452(08)70260-0.

Wicha, N.Y., Dickson, D.S., Martinez-Lincoln, A., 2018. Arithmetic in the bilingual brain. Language and Culture in Mathematical Cognition, pp. 145–172. https://doi.org/10.1016/b978-0-12-812574-8.00007-9.

Woodcock, R.W., McGrew, K.S., Mather, N., 2001. Woodcock-Johnson III Tests of Cognitive Abilities. Riverside Publishing Company.

Xuan, D., Wang, S., Yang, Y., Meng, P., Xu, F., Yang, W., et al., 2007. Age difference in numeral recognition and calculation: an event-related potential study. Child Neuropsychol. 13 (1), 1–17. https://doi.org/10.1080/09297040600760465.

Zhou, X., Booth, J.R., Lu, J., Zhao, H., Butterworth, B., Chen, C., Dong, Q., 2011. Age-independent and age-dependent neural substrate for single-digit multiplication and addition arithmetic problems. Dev. Neuropsychol. 36 (3), 338–352. https://doi.org/10.1080/87565641.2010.549873.