The influence of cardiorespiratory fitness on strategic, behavioral, and electrophysiological indices of arithmetic cognition in preadolescent children

R. Davis Moore , Eric S. Drollette , Mark R. Scudder , Aashiv Bharij and Charles H. Hillman *

Neurocognitive Kinesiology Laboratory, Kinesiology, University of Illinois, Urbana, IL, USA

The current study investigated the influence of cardiorespiratory fitness on arithmetic cognition in forty 9–10 year old children. Measures included a standardized mathematics achievement test to assess conceptual and computational knowledge, self-reported strategy selection, and an experimental arithmetic verification task (including small and large addition problems), which afforded the measurement of event-related brain potentials (ERPs). No differences in math achievement were observed as a function of fitness level, but all children performed better on math concepts relative to math computation. Higher fit children reported using retrieval more often to solve large arithmetic problems, relative to lower fit children. During the arithmetic verification task, higher fit children exhibited superior performance for large problems, as evidenced by greater d' scores, while all children exhibited decreased accuracy and longer reaction time for large relative to small problems, and incorrect relative to correct solutions. On the electrophysiological level, modulations of early (P1, N170) and late ERP components (P3, N400) were observed as a function of problem size and solution correctness. Higher fit children exhibited selective modulations for N170, P3, and N400 amplitude relative to lower fit children, suggesting that fitness influences symbolic encoding, attentional resource allocation and semantic processing during arithmetic tasks. The current study contributes to the fitness-cognition literature by demonstrating that the benefits of cardiorespiratory fitness extend to arithmetic cognition, which has important implications for the educational environment and the context of learning.

Keywords: addition, pediatric-cognition, strategy, ERPs, mathematics

INTRODUCTION

Recent research suggests that cardiorespiratory fitness and physical activity (PA) are positively associated with neurocognitive health across the lifespan (Colcombe et al., 2004a,b; Hillman et al., 2005, 2006; Kramer et al., 2006; Pontifex et al., 2009; Smith et al., 2010; Erickson et al., 2011; see Hillman et al., 2008 for review), but the majority of research has focused on adult populations with fewer efforts directed toward understanding the relation of cardiorespiratory fitness and PA to neurocognition during development. As children have become increasingly sedentary and opportunities for PA during the school day have diminished (Institute of Medicine of the National Academies, 2013), illuminating the neurocognitive benefits resulting from cardiorespiratory fitness and PA have never been more important. What research exists indicates that cardiorespiratory fitness and PA are also positively associated with neurocognition during development, with disproportionate benefits witnessed on the behavioral and neural levels for tasks requiring variable amounts of attention and cognitive control (Hillman et al., 2005, 2009; Buck et al., 2008; Chaddock et al., 2011; Pontifex et al., 2011; Voss et al., 2011; Moore et al., 2013). However, the specificity of the relation between cardiorespiratory fitness and PA in developing populations continues to unfold (Tomporowski, 2003; Sibley and Etnier, 2003; Castelli et al., 2007; Buck et al., 2008; Hillman et al., 2009; Pontifex et al., 2011; Moore et al., 2013).

One area receiving increasing attention is the relation of cardiorespiratory fitness to academic achievement. Both larger-scale cross-sectional (California Department of Education, 2001, 2005; Cottrell et al., 2007; Chomitz et al., 2009), and smaller-scale experimental studies (Castelli et al., 2007; Wittberg et al., 2012) have found a positive relation of fitness to linguistic and arithmetic indices of academic achievement. Arithmetic achievement is of particular interest given that arithmetic cognition is a fundamental skill in modern society, plays an important role in everyday life (Rips et al., 2008; Chen et al., 2013) and is a critical skill set for children to master (El Yagoubi et al., 2005; Menon, 2010). Recently, research efforts have been directed toward understanding the development of arithmetic proficiency on both the behavioral and neural level to understand how this skill set is acquired and effectively maintained across the lifespan (Rips et al., 2008; Imbo and Vandierendonck, 2008; Chen et al., 2013). While several demographic and health factors have been found to mediate arithmetic development and achievement (White, 1982; Geary et al., 2004; Sirin, 2005; Castelli et al., 2007; Chomitz et al., 2009), in general, the development of arithmetic proficiency is characterized by a shift in strategy selection from effortful, inefficient

strategies to more automated and efficient strategies (Siegler, 1986). Thus, arithmetic proficiency is contingent on both strategy selection and strategy efficiency (Imbo and Vandierendonck, 2008).

Strategy selection refers to the procedure necessary to solve a problem, and strategy efficiency refers to the speed and accuracy at which a solution is produced or verified (Imbo and Vandierendonck, 2008). Children typically rely on one of three strategies to solve arithmetic problems: (1) finger and verbal counting, which are effortful and less efficient strategies used during initial learning, (2) decomposition (i.e., $8 + 7 = 5 + 3 + 5 + 2$), and (3) retrieval. These last two strategies are more automated and efficient, and are characteristic of increasing arithmetic skill (Ashcraft, 1982; Siegler, 1986; Roussel et al., 2002; Imbo and Vandierendonck, 2008; Cho et al., 2011). Accordingly, the developmental shift from finger and verbal counting to decomposition and retrieval strategies leads to quicker and more accurate solution production and verification (Geary et al., 2004; Imbo and Vandierendonck, 2008). This shift in strategy is most evident in the second and third grades (Ashcraft and Fierman, 1982; Geary et al., 1987, 2004), and is contingent on the development of children's conceptual understanding of counting (Siegler, 1987; Geary et al., 2004), phonological abilities (De Smedt et al., 2010), and the development of semantic memory networks between problem stems and solutions (Siegler and Shrager, 1984; Cho et al., 2011).

In addition to standardized achievement tests, the arithmetic verification task has been of particular utility for revealing behavioral and neural processes associated with arithmetic calculation across the lifespan (Niedeggen et al., 1999; El Yagoubi et al., 2003; Galfano et al., 2004; Jost et al., 2004; Núñez-Peña et al., 2006, 2011; Xuan et al., 2007; Imbo and Vandierendonck, 2008; De Smedt et al., 2010; Prieto-Corona et al., 2010). During arithmetic verification tasks, individuals are presented with problems in the form of $a + b = c$, and must verify whether the solution is correct or incorrect. On the behavioral level, solution verification has been characterized by longer RT and decreased accuracy (ACC) for incorrect relative to correct solutions (Niedeggen and Rosler, 1999; Campbell and Fugelsang, 2001; Domahs and Delazer, 2005; Jasinski and Coch, 2012); a phenomenon known as the split effect. Solution verification has also been characterized by longer RT and decreased ACC for large (>10) relative to small (<10) solutions (Groen and Parkman, 1972; Zbrodoff and Logan, 2005; Imbo and Vandierendonck, 2008; Núñez-Peña et al., 2011); a phenomenon known as the problem size effect. Thus, verification tasks enable the evaluation of arithmetic processes across multiple dimensions of difficulty (i.e., correctness, size).

Electroencephalography (EEG) and event-related potential in particular (ERPs) have proven to be an invaluable tool for evaluating the neural underpinnings of arithmetic cognition (El Yagoubi et al., 2005; Muluh, 2011; Jasinski and Coch, 2012). During arithmetic verification, ERPs time-locked to solution presentation reliably reveal a P3, N400-like negativity, and a late positive component (LPC) in adults. The arithmetic P3 is larger for correct relative to incorrect solutions (Niedeggen et al., 1999; Galfano et al., 2004; Jost et al., 2004; Núñez-Peña et al., 2011; Jasinski and Coch, 2012) and has been linked to the classic P3b, (Niedeggen

et al., 1999; Jost et al., 2004). The arithmetic N400 is larger for incorrect, relative to correct solutions (Niedeggen et al., 1999; Jost et al., 2004; Prieto-Corona et al., 2010; Jasinski and Coch, 2012), and has been linked to the N400 observed in other paradigms, suggesting that it is an index of semantic information processing (Kutas and Federmeier, 2000, 2011; Federmeier and Laszlo, 2009). The LPC is larger for incorrect relative to correct solutions and is hypothesized to be an index of plausibility processing (i.e., given $a + b$, is solution c reasonable?; Niedeggen et al., 1999; Jost et al., 2004; Domahs et al., 2007; Jasinski and Coch, 2012); linking this component to the P600 (Núñez-Peña and Honrubia-Serrano, 2004; Núñez-Peña et al., 2004). In addition, earlier ERP components such as the N1/N170 have been systematically modulated during numerical paradigms (Dehaene, 1996; Szűcs and Goswami, 2007; Hyde and Spelke, 2009, 2012; Palomares et al., 2011); however, the functional interpretation of these components remains controversial (Feigenson et al., 2004; Muluh, 2011; Heine et al., 2012) and seldom explored during arithmetic verification tasks (He et al., 2011; Muluh et al., 2011).

Despite numerous investigations examining the electrophysiological processes underlying arithmetic verification in adults, a paucity of data exists for children with only a few initial studies comparing children and adults (Xuan et al., 2007; Prieto-Corona et al., 2010). For example, Prieto-Corona et al. (2010) compared 8–10 year old children and young adults during a multiplication-verification task. In addition to longer RT and decreased ACC, the children exhibited larger N400 amplitude and longer N400 latency for incorrect solutions relative to adults. Further, adults, but not children, displayed a LPC during incorrect solution presentation. Thus, in addition to behavioral differences, children also quantitatively and qualitatively differ from adults on the electrophysiological level during arithmetic performance. As such, additional research is warranted to detail the neurodevelopmental shifts that give rise to mature arithmetic cognition, as well as the potential health factors, which may mediate this development.

The current study evaluated arithmetic performance in higher and lower fit children by employing both a standardized achievement test as well as an experimental addition-verification task, which consisted of small (<10) and large (>10) solutions, and afforded the measurement of electrophysiological activity. Furthermore, to assess strategy selection, participants were asked to report how they solved small and large addition problems, which appeared during both the standardized achievement assessment and experimental task. Irrespective of fitness, all children were expected to demonstrate longer RT and decreased ACC for incorrect relative to correct solutions, irrespective of solution size. It was also predicted that all children would demonstrate longer RT and decreased ACC for large relative to small solutions, irrespective of solution correctness; thus replicating prior work (Imbo and Vandierendonck, 2008; Prieto-Corona et al., 2010; Cho et al., 2011). Children were further expected to exhibit larger P3 amplitude for correct relative to incorrect solutions and larger N400 amplitude for incorrect relative to correct solutions. Based on prior work (Prieto-Corona et al., 2010), children were not expected to exhibit a LPC, indicative of a protracted development in plausibility processing.

With respect to fitness, higher fit children were expected to demonstrate superior performance for standardized math achievement and report more frequent use of retrieval than their lower fit counterparts. It was further expected that higher fit relative to lower fit children would demonstrate differences in performance on the behavioral and electrophysiological levels during the arithmetic verification task. Specifically, higher fit children were expected to respond more quickly and accurately during incorrect solutions across problem sizes, and this effect would be selectively greater for large problems. In addition, higher fit relative to lower fit children were predicted to demonstrate more flexible deployment of attention, as indexed by smaller P3 amplitude during small problem solutions and larger P3 amplitude during large problem solutions. Lastly, we predicted that higher fit children would demonstrate larger N400 amplitude during incorrect problem solutions indicating facilitated semantic access for discriminating between incorrect and correct solutions.

MATERIALS AND METHODS
PARTICIPANT CHARACTERISTICS
Forty preadolescent children aged 9–10, (16 female) were recruited from the East-Central Illinois region. Participants were bifurcated into higher (>70th percentile) or lower (<30th percentile) fitness groups based on age-specific norms (Shvartz and Reibold, 1990). Maximal aerobic capacity (VO$_{2max}$) was based on the volume of oxygen consumed during maximum capacity exercise (ml/kg·min^{-1}). **Table 1** lists demographic and fitness information for the sample. No child received special education services related to mental or physical disabilities and all participants and their legal guardians provided written informed assent/consent in accordance with the Institutional Review Board at the University of Illinois.

Prior to testing, legal guardians completed a health history and demographics questionnaire, indicating that their child was

Table 1 | Participant demographics data for higher and lower fit children.

Measure	Higher fit	Lower fit
Age (years)	9.9 (0.7)	10.1 (0.6)
Gender (M/F)	13/7	11/9
Grade	4.0 (0.8)	4.3 (0.6)
SES	2.0 (0.8)	2.4 (0.7)
Tanner	1.2 (0.3)	1.3 (0.4)
K-BIT	120.8 (11.5)	119.9 (11.8)
BMI percentile (%)	35.6 (28.1)	52.7 (32.0)
BMI	16.9 (3.5)	19.1 (4.7)
Vo2 percentile (%)*	82.7 (7.1)	28.1 (7.9)
Vo2 relative	52.7 (5.1)	41.43 (4.2)
Computation percentile (%)	76.8 (23.9)	77.8 (21.7)
Concepts percentile (%)	87.0 (12.1)	89.5 (17.0)
Composite percentile (%)	88.1 (13.8)	87.5 (16.1)

*Tanner refers to the Tanner pubertal timing scale; SES, socio-economic status; K-BIT, Kaufmann Brief Intelligence Test; BMI, body mass index; VO2 refers to aerobic fitness; Computation, concepts, and composite refer to the sub-sections and combined composite score of the KTEA-2 achievement test. *p < 0.05.*

free of neurological diseases or physical disabilities. The Kaufman Brief Intelligence Test 2 (KBIT-2; Kaufman and Kaufman, 2004) was administered to each participant to create a composite intelligence quotient (IQ). The Attention-Deficit Hyperactivity Disorder Rating Scale IV (DuPaul et al., 1998) was completed by guardians to screen for the presence of attentional disorders (as indexed by scores above 14 and 22 for females and males, respectively). In cooperation with the child, guardians completed a modified Tanner Staging System (Taylor et al., 2001) to assess pubertal timing. Subsequently, all participants were at or below a score of 2 (i.e., prepubescent) at time of testing. In addition, SES was assessed by computing a trichotomous index based on three variables: (a) participation in a free or reduced-price lunch program at school; (b) the highest level of education obtained by the mother and father; and (c) number of parents who worked full time (Birnbaum et al., 2002). Lastly, all participants demonstrated right-handedness as measured by the Edinburgh Handedness Inventory (Oldfield, 1971).

CARDIORESPIRATORY FITNESS ASSESSMENT
VO$_{2max}$ was measured on a motor–driven treadmill using a modified Balke protocol, which is recommended for graded exercise testing with children (American College of Sports Medicine, 2010). Prior to testing, participants had their height and weight measured, were fitted with a Polar heart rate (HR) monitor (Polar Wear Link® + 31, Polar Electro, Finland), and underwent a brief warm-up period. The treadmill was then set to a constant speed during the test, while grade increments of 2.5% occurred every 2 min until volitional exhaustion. Oxygen consumption was measured using a computerized indirect calorimetry system (ParvoMedics True Max 2400) with averages for oxygen uptake and respiratory exchange ratio (RER) assessed every 20 s. Concurrently, ratings of perceived exertion (RPE) were measured every 2 min using the children's OMNI scale (Utter et al., 2002). VO$_{2max}$ was established when children met a minimum of 2 of the following 4 criteria: (1) a plateau in oxygen uptake corresponding to an increase of less than 2 ml/kg·min^{-1} despite an increase in exercise workload; (2) a peak HR ≥185 beats per minute (bpm; American College of Sports Medicine, 2010) and a HR plateau (Freedson and Goodman, 1993); (3) RER ≥1.0 (Bar-Or, 1983); and/or (4) ratings on the children's OMNI scale of perceived exertion ≥8 (Utter et al., 2002). Relative peak oxygen consumption was expressed in milliliters of oxygen consumed per kilogram of body weight per minute.

EEG RECORDING
Electroencephalographic (EEG) activity was recorded from 64 sintered 10 mm Ag-AgCl electrodes (FPz, Fz, FCz, Cz, CPz, Pz, POz, Oz, FP1/2, F7/5/3/1/2/4/6/8, FT7/8, FC3/1/2/4, T7/8, C5/3/1/2/4/6, M1/2, TP7/8, CB1/2, P7/5/3/1/2/4/6/8, PO7/5/3/4/6/8, O1/2), arranged according to the International 10-10 system (Chatrian et al., 1985) using a Neuroscan Quik-cap (Compumedics, Inc, Charlotte, NC). EEG activity was referenced to averaged mastoids (M1, M2), with AFz serving as the ground electrode. Impedance was kept below 10 kΩ. Additional electrodes were placed above and below the left orbit and on the outer canthus of each eye to monitor electro-oculographic (EOG)

activity with a bipolar recording. Continuous raw EEG data were collected using Neuroscan Scan software (v 4.5) and amplified through a Neuroscan Synamps 2 amplifier with a 24 bit A/D converter and ± 200 millivolt (μV) input range (763 μV/bit resolution). Data were sampled at a rate of 500 Hz and amplified 500 times with a DC to 70 Hz filter, and a 60 Hz notch filter.

TASKS
Achievement
Participants were administered the mathematics subsections of the Kaufman Test of Academic and Educational Achievement 2 (KTEA-2; Kaufman and Kaufman, 2004), which included tests of math concepts and computation. The subtest begins by testing concepts such as cardinality, ordinality, comparing quantities, as well as basic arithmetic and rounding. As problems increase in difficulty, algebraic, calculus, and trigonometry concepts are required. Participants were given a scratch paper and a pencil, but were not allowed to use a calculator. The math computation subsection is a 72-item subtest, which begins with basic arithmetic operations including: adding, subtracting, multiplying, and dividing whole numbers of increasing magnitude, as well as fractions. Later problems require calculations involving exponents, decimals, negatives, and unknown variables. Again, participants were provided with scratch paper and a pencil, but were not allowed to use a calculator. Participants' scores were entered into the normative age database to provide an achievement percentile score for each subtest as well as composite match achievement percentile score.

Arithmetic verification task
The current arithmetic verification task was modeled on parameters provided by Núñez-Peña and Suárez-Pellicioni (2012). However, given the younger age of children in the current study and preliminary pilot testing, the largest problem combinations from Núñez-Peña's paradigm were not employed. All problems were expressed in the form of $a + b = c$. For each problem two operand orders were created ($a + b = c$, $b + a = c$). Small problems used single-digit operands between 1 and 4 and large problems used single-digit operands between 6 and 9. Ties (e.g., $3 + 3$), and consecutive even operands (e.g., $2 + 4$) were excluded, and the solution was never the product of $a \times b$. For each problem and operand order, both a correct and incorrect solution were created with incorrect solutions being either lesser or greater by 1 than the correct solution. Thus, all incorrect solutions were small split, and parity was controlled.

Each trial consisted of stimuli presented sequentially in the following order: a fixation dot presented for 500 ms, the first operand presented for 1000 ms, a "+" sign presented for 500 ms, the second operand presented for 2000 ms, and then the solution, which was surrounded by a box and remained on the screen until the participant responded or a maximum of 2000 ms elapsed. The inter-stimulus interval was 100 ms and participants were instructed to respond as quickly and accurately as possible. Participants were counterbalanced according to correct response selection, with half of the participants instructed to make a right hand thumb press on a response pad if the solution was correct and the other half instructed to make a left thumb press if the solution was correct. Response assignments were further counterbalanced across fitness groupings. Participants completed two blocks of small problems and two blocks of large problems, which were counterbalanced across participants. Thus, all participants completed 240 trials, 120 for each problem set size, with 60 correct and 60 incorrect solutions presented randomly for each problem set size (see **Figure 1**).

LABORATORY PROCEDURE
Day 1
Participants and their guardians completed an informed assent and informed consent, respectively. Next, participants completed the Edinburgh Handedness Inventory followed the KBIT-2, which was administered by a trained experimenter. Participants then completed the mathematics portion of the KTEA-2. Concurrently, participants' legal guardians completed the health history and demographics questionnaire, the ADHD Rating Scale IV, the modified Tanner Staging System, and the Physical Activity Readiness Questionnaire (Thomas et al., 1992). Participants then had their height and weight measured and completed the cardiorespiratory fitness assessment. Upon completion, participants were afforded a cool down period and remained in the laboratory until their HR returned to within 10 beats per minute of their resting HR.

Day 2
Participants returned to the laboratory and were outfitted with an EEG cap before being seated in an electrically and acoustically attenuated testing chamber. Following the provision of instructions for the arithmetic verification task, participants were given the opportunity to ask questions, and then performed a practice block of 30 trials prior to each problem set size. The experimenter observed participants during the practice trials and checked their performance to ensure that they understood the task. If a participant's task performance was below 60%, another practice block was administered. Upon the completion of the task, participants
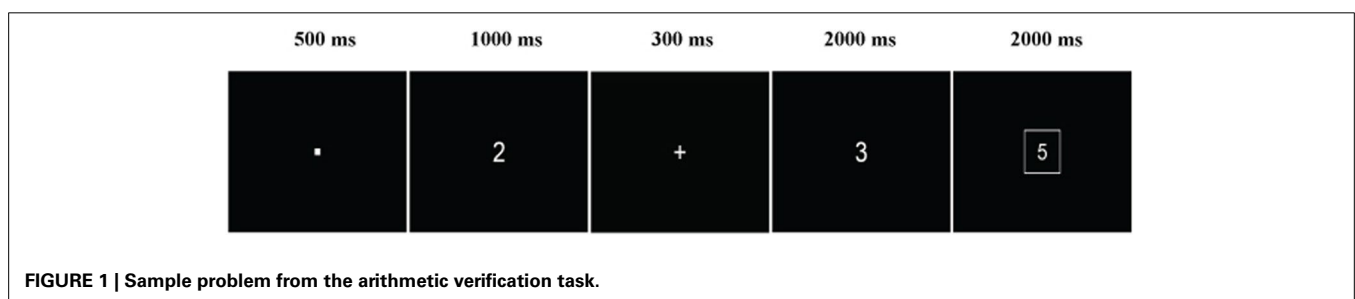
FIGURE 1 | Sample problem from the arithmetic verification task.

were briefed on the purpose of the experiment, and received $10/h remuneration.

BEHAVIORAL DATA REDUCTION

Strategy

Children were asked to report how they solved a small and large addition problem during the computation portion of KTEA-2 achievement test. Similar to previous studies (Geary et al., 2004), children were asked "can you tell me how you got the answer?" and based on the child's response and experimenter's observation, responses were classified into three categories: counting (finger/verbal), decomposition ($4 + 7 = 4 + 5 + 2$), or retrieval ("just knew it"). Responses were coded as 1 for counting, 2 for decomposition, and 3 for retrieval. Thus, each participant received a score of 1, 2, or 3 per problem.

Mathematics achievement

A trained experimenter graded children's responses such that children received a 1 for each correct response and a 0 for an incorrect response. Scores were then tallied to generate a total score for each sub-section and entered into a normative database of values. Thus, each child received an age-normed achievement percentile for each sub-section, as well as a composite achievement percentile score.

Arithmetic verification task

Behavioral data were collected in terms of RT (time in milliseconds from stimulus presentation until manual response) for correct trials, and ACC (percentage of correct responses) for each task condition. In accord with previous research (Geary, 2010; Núñez-Peña and Suárez-Pellicioni, 2012), d' [z (hit rate) - z (false alarm rate)] scores were calculated for each problem size.

ELECTROPHYSIOLOGICAL DATA REDUCTION

Prior to averaging, an off-line EOG reduction procedure was applied to individual trials via a spatial filter (Compumedics Neuroscan, 2003), which performed a principle component analysis (PCA) to determine the major components that characterize the EOG artifact between all channels. This procedure then reconstructed the original channels without the artifact components (Compumedics Inc, Neuroscan, 2003). Trials with a response error or artifact exceeding $\pm75\,\mu$V were rejected and artifact free data were retained for averaging. An average of 43 ($\pm$ 2) trials and 42 ($\pm$ 3) trials were retained for large-correct and large-incorrect solutions respectively, and 48 ($\pm$ 1) trials and 44 ($\pm$ 2) trials were retained for small-correct and small-incorrect solutions, respectively. Higher and lower fit participants did not differ in the number of trials retained for averaging, p's $\geq$ 0.83.

Stimulus-locked components were created using epochs from -100 to 1000 ms around solution stimuli and were baseline corrected using the 100-ms pre-stimulus period. Data were filtered with a zero phase shift 30-Hz low-pass cutoff (24 dB/octave rolloff). The P1 component was identified as the mean amplitude within a 30 ms interval surrounding the largest positive-going peak within 75–150 ms latency. The N170 component was identified as the mean amplitude within a 30 ms interval surrounding the largest negative-going peak within 100–200 ms latency. The P3 component was identified as the mean amplitude within a

50 ms interval surrounding the largest positive-going peak within 300–600 ms latency. The N400 component was identified as the mean amplitude within a 50 ms interval surrounding largest negative-going peak within 300–500 ms latency. Amplitude was measured as the difference between the mean pre-stimulus baseline and mean peak-interval amplitude; peak latency was defined as the time point corresponding to the maximum local peak amplitude.

STATISTICAL ANALYSIS

Statistical analyses were performed using SPSS version 19.0 (SPSS Inc., Chicago, IL) and statistical significance was noted when $p < 0.05$. Paired sample and independent samples t-tests were conducted to evaluate both academic achievement scores and strategy reports. Behavioral data were analyzed using a 2 (Group: higher fit, lower fit) $\times$ 2 (Correctness: correct, incorrect) $\times$ 2 (Problem Size: small, large) repeated-measures ANOVA for the arithmetic verification task, with fitness group entered as a between-subjects factor. In addition, d' scores for the arithmetic verification task were analyzed using a 2 (Group: higher fit, lower fit) $\times$ 2 (Problem Size: small, large) repeated-measures ANOVA. All ANOVAs used the Greenhouse-Geisser correction to correct for violations of sphericity and Bonferroni corrected t-tests were utilized to evaluate post-hoc significance.

Electrophysiological analysis was conducted separately on P1, N170, P3, and N400 component values (i.e., amplitude, latency). Similar to previous investigations (Prieto-Corona et al., 2010; Muluh et al., 2011; Núñez-Peña and Suárez-Pellicioni, 2012) regions of interest (ROIs) were created. Specifically, P1, N170, and P3 component values were formulated by averaging electrode sites into 3 regions: left (P7, PO7, P5, PO5, P3, PO3), center (P1, PZ, POZ, P2), and right (P8, PO8, P6, PO6, P4, PO4) using similar factorial models as described above with the addition of a region factor. Based on difference waves, N400 component values for each participant were analyzed by decomposing electrode sites into 2 ROI's: left (C5, CP5, P5, C3, CP3, P3, C1, CP1, P1) and right (C6, CP6, P6, C4, CP4, P4, C2, CP2, P2) and were submitted to similar factorial models as described above, with the addition of a region factor.

RESULTS

BEHAVIOR

Mathematics achievement

Achievement data are reported in **Table 1**. Analysis of achievement data revealed that all participants' scored significantly higher on the math concepts relative to the math computation section of the achievement test, [$t_{(39)} = 3.84$, $p < 0.01$]. No fitness group differences were realized for math computation, concepts, or composite achievement percentile, [t's$_{(38)} \leq 0.36$, p's $\geq$ 0.72].

Strategy

Analysis of strategy revealed a main effect of problem size, indicating that all participants reported relying more on retrieval than procedural strategies (counting, decomposition) for small ($m = 2.9 \pm 0.5$) relative to large problems ($m = 2.5 \pm 0.7$), [$F_{(1, 38)} = 10.50$, $p < 0.01$, $\eta^2 = 0.21$]. However, this effect, was superseded

by problem size × fitness interaction, [$F_{(1, 38)} = 5.65$, $p = 0.02$, $\eta^2 = 0.13$]. *Post-hoc* analysis revealed a significant trend indicating that higher fit children reported relying on retrieval more frequently than lower fit children during large problem solutions, [$t_{(39)} = 2.30$, $p < 0.03$].

Arithmetic verification task

RT. Analysis revealed effects of problem size, [$F_{(1, 38)} = 19.90$, $p < 0.01$, $\eta^2 = 0.34$], and correctness, [$F_{(1, 38)} = 89.31$, $p < 0.01$, $\eta^2 = 0.70$], indicating that all participants responded more quickly to small ($m = 863.58 \pm 158.8$), relative to large ($m = 932.84 \pm 166.6$) problems, and for correct ($m = 838.82 \pm 170.9$) relative to incorrect ($m = 961.8 \pm 170.9$) problems. Analysis did not reveal any significant effects of fitness, [F's$_{(1, 38)} \leq 1.89$, $p \geq 0.18$, η's$^2 \leq 0.05$].

ACC. Analysis revealed effects of problem size, [$F_{(1, 38)} = 23.64$, $p < 0.01$, $\eta^2 = 0.38$], and correctness, [$F_{(1, 38)} = 7.92$, $p < 0.01$, $\eta^2 = 0.17$], which were superseded by a problem size × correctness interaction, [$F_{(1, 38)} = 5.71$, $p = 0.02$, $\eta^2 = 0.13$]. *Post-hoc* analysis revealed that all participants responded less accurately for large-incorrect problems ($m = 74.3 \pm 15.9$), relative to small-correct ($m = 84.7 \pm 8.4$), small-incorrect ($m = 81.7 \pm 10.8$), and large-correct ($m = 81.2 \pm 12.7$) problems, [t's$_{(38)} \geq 3.15$, $p < 0.01$]. No effect of fitness was observed, [F's$_{(1, 38)} \leq 2.42$, p's ≥ 0.13, $\eta2'$s ≤ 0.06].

d′. Analysis revealed a main effect of problem size, [$F_{(39, 1)} = 5.94$, $p = 0.02$, $\eta^2 = 0.14$], indicating that all participants were more accurate at detecting correct (and rejecting incorrect) solutions for small ($m = 2.7 \pm 0.5$) relative to large ($m = 2.4 \pm 0.7$) problems. Analysis also revealed a fitness × problem size interaction, [$F_{(1, 38)} = 5.0$, $p = 0.04$, $\eta^2 = 0.12$]. *Post-hoc* analysis revealed that higher fit ($d = 2.7 \pm 0.5$) relative to lower fit ($d = 2.2 \pm 0.8$) children more accurately detected correct and incorrect solutions only for large size problems, [$t_{(38)} = 2.4$, $p = 0.02$].

ELECTROPHYSIOLOGICAL DATA
P1
Amplitude and latency data for the P1 component are presented in **Table 2**. Omnibus analysis of amplitude revealed a main effect of region, [$F_{(1, 38)} = 13.72$, $p < 0.01$, $\eta^2 = 0.27$], which was superseded by a problem size × region interaction, [$F_{(1, 38)} = 36.01$, $p < 0.01$, $\eta^2 = 0.49$]. Analysis also revealed a problem size × correctness interaction, [$F_{(1, 38)} = 23.40$, $p < 0.01$, $\eta^2 = 0.38$], and correctness × region interaction, [$F_{(1, 38)} = 7.21$, $p < 0.01$, $\eta^2 = 0.16$], which were superseded by a problem size × correctness × region interaction, [$F_{(1, 38)} = 12.36$, $p < 0.01$, $\eta^2 = 0.25$]. *Post-hoc* analysis of the 3-way interaction revealed that P1 amplitude for all participants was greater during small-correct problems over the right ROI ($m = 9.2 \pm 7.6$) relative to the center ROI ($m = 6.2 \pm 4.6$), and during small-incorrect problems over the right ROI ($m = 9.7 \pm 8.2$) relative to the center ROI ($m = 6.5 \pm 4.8$). Further, amplitude during large problems was greater over the right ROI ($m = 8.0 \pm 4.6$) than the center ($m = 3.9 \pm 3.2$) and left ($m = 5.7 \pm 4.2$) ROIs, and amplitude was greater over the left ROI ($m = 5.7 \pm 4.2$) than the

Table 2 | Amplitude and latency data for the N170 and P1 components for higher fit and lower fit children.

PI amplitude (μv)	Higher fit	Lower fit
Small-correct-left	8.0 (6.2)	11.2 (7.3)
Small-incorrect-left	8.4 (5.8)	12.1 (6.6)
Large-correct-left	7.7 (5.0)	8.2 (5.5)
Large-incorrect-left	7.1 (4.8)	8.3 (5.5)
Small-correct-center	4.6 (2.7)	8.0 (5.4)
Small-incorrect-center	4.6 (2.6)	6.6 (5.5)
Large-correct-center	3.0 (1.9)	5.5 (3.7)
Large-incorrect-center	2.2 (2.0)	5.1 (4.1)
Small-correct-right	7.5 (5.2)	8.9 (7.6)
Small-incorrect-right	10.4 (7.2)	14.2 (11.8)
Large-correct-right	10.5 (7.2)	10.5 (8.7)
Large-incorrect-right	9.7 (6.0)	10.3 (8.9)
P1 latency (ms)		
Small-correct-left	121.0 (20.3)	123.0 (17.0)
Small-incorrect-left	131.7 (19.2)	121.6 (19.3)
Large-correct-left	118.2 (18.7)	123.4 (16.4)
Large-incorrect-left	118.6 (18.4)	124.8 (16.5)
Small-correct-center	137.4 (22.1)	140.8 (22.1)
Small-incorrect-center	137.4 (23.8)	139.5 (17.0)
Large-correct-center	133.3 (19.3)	142.6 (19.6)
Large-incorrect-center	135.6 (29.3)	141.2 (27.7)
Small-correct-right	120.1 (18.9)	120.1 (18.9)
Small-incorrect-right	119.2 (16.6)	126.5 (21.5)
Large-correct-right	119.4 (21.2)	117.6 (18.9)
Large-incorrect-right	117.2 (23.4)	123.4 (21.1)
N170 amplitude (μv)		
Small-correct-left	−6.1 (6.3)	−3.7 (5.4)
Small-incorrect-left	−5.7 (6.9)	−2.9 (7.6)
Large-correct-left	−6.8 (4.9)	−4.5 (6.0)
Large-incorrect-left	−6.2 (4.1)	−3.5 (4.6)
Small-correct-center	−3.3 (4.6)	−2.3 (4.7)
Small-incorrect-center	−2.7 (4.5)	−2.3 (4.9)
Large-correct-center	−2.6 (2.8)	−2.1 (4.1)
Large-incorrect-center	−3.4 (2.7)	−2.3 (4.0)
Small-correct-right	−3.6 (4.5)	−0.4 (2.7)
Small-incorrect-right	−5.5 (5.5)	0.03 (6.5)
Large-correct-right	−5.4 (6.5)	−2.8 (3.8)
Large-incorrect-right	−5.4 (6.3)	−1.5 (4.0)
N170 latency (ms)		
Small-correct-left	195.7 (18.6)	193.6 (23.7)
Small-incorrect-left	193.4 (16.7)	192.1 (22.9)
Large-correct-left	200.8 (19.7)	200.1 (19.6)
Large-incorrect-left	200.1 (15.9)	197.9 (23.2)
Small-correct-center	201.8 (17.4)	202.8 (13.1)
Small-incorrect-center	198.8 (19.0)	200.5 (13.8)
Large-correct-center	201.5 (20.8)	204.9 (16.3)
Large-incorrect-center	208.2 (18.8)	206.4 (17.9)
Small-correct-right	198.5 (13.4)	198.4 (21.0)
Small-incorrect-right	196.1 (15.5)	201.6 (20.1)
Large-correct-right	200.0 (16.7)	194.7 (21.9)
Large-incorrect-right	202.9 (17.8)	195.2 (23.2)

μv, microvolts; ms, milliseconds.

center ROI ($m = 3.9 \pm 3.2$), [$t's_{(38)} \geq 2.65$, $p's \leq 0.01$]. In addition, amplitude during small-correct problems was greater ($m = 6.2 \pm 4.5$) than for large-correct problems ($m = 4.2 \pm 3.2$) over the center ROI, [$t_{(39)} = 3.03$, $p < 0.01$], and amplitude during small-incorrect problems was greater than large-incorrect problems over the left (small: $m = 7.7 \pm 5.1$; large: $m = 5.6 \pm 4.2$) and center (small: $m = 6.5 \pm 4.8$; large: $m = 3.5 \pm 3.6$) ROIs.

In addition, analysis revealed a fitness x correctness interaction, [$F_{(1, 38)} = 3.9$, $p = 0.05$, $\eta^2 = 0.09$], suggesting that lower relative to higher fit children exhibited larger P1 amplitude during incorrect problem solutions. However, *post-hoc* tests failed to reveal significant effects upon decomposition of the interaction, [$t's_{(38)} \leq 1.87$ $p's \geq 0.07$].

Analysis of latency revealed a main effect of region, [$F_{(1, 38)} = 18.78$, $p < 0.01$, $\eta^2 = 0.33$], which was superseded by a region × correctness interaction, [$F_{(1, 38)} = 32.64$, $p < 0.01$, $\eta^2 = 0.46$]. *Post-hoc* tests revealed longer latency over the center ROI, for correct and incorrect solutions (correct: $m = 138.5 \pm 19.2$; incorrect: $m = 138.4 \pm 20.7$), relative to the right (correct: $m = 119.1 \pm 15.4$; incorrect: $m = 121.8 \pm 17.3$) and left (correct: $m = 121.5 \pm 13.1$; incorrect: $m = 124.2 \pm 13.6$) ROIs, [$t's_{(39)} \geq 5.32$, $p's < 0.01$]. Analysis did not reveal any effect of fitness, [$F's_{(1, 38)} \leq 1.23$, $p's \geq 0.27$, $\eta's^2 \leq 0.03$].

N170

Amplitude and latency data for the N170 are presented in **Table 2**. Analysis of amplitude revealed a main effect of region, [$F_{(1, 38)} = 6.22$, $p = 0.02$, $\eta^2 = 0.14$]. *Post-hoc* analysis indicated that all participants demonstrated greater amplitude over the left ($m = -4.9 \pm 5.3$) relative to the center ($m = -2.6 \pm 3.8$) and right ($m = -3.1 \pm 4.8$) ROIs, [$t's_{(39)} \geq 2.5$ $p's < 0.02$]. Further, a main effect of fitness was revealed, [$F_{(1, 38)} = 5.63$, $p = 0.02$, $\eta^2 = 0.13$], which was superseded by a fitness × correctness interaction, [$F_{(1, 38)} = 4.61$, $p = 0.03$, $\eta^2 = 0.11$. *Post-hoc* tests revealed that higher fit ($m = -5.7 \pm 4.5$) relative to lower fit ($m = -1.9 \pm 4.3$) children demonstrated larger N170 amplitude only during incorrect problem verification, [$t_{(38)} = 2.66$, $p = 0.01$], while no such differences were observed for correct problem verification, [$t_{(38)} = 1.97$, $p = 0.06$]. No effects of fitness, problem size, correctness or region were observed for N170 latency, [$F's_{(1, 38)} \leq 1.09$, $p's \geq 0.30$].

P3

P3 amplitude and latency data are presented in **Table 3**. Analysis of amplitude revealed a main effect of problem size, [$F_{(1, 38)} = 15.30$, $p = 0.01$, $\eta^2 = 2.87$], which was superseded by a problem size × correctness × region interaction, [$F_{(1, 38)} = 5.23$, $p = 0.01$, $\eta^2 = 1.21$]. *Post-hoc* tests revealed that all participants demonstrated greater P3 amplitude over the right ROI during small problems ($m = 10.7 \pm 6.1$), relative to the left ($m = 7.5 \pm 4.5$) and center ($m = 7.3 \pm 4.3$), ROIs during large problems, [$t's_{(39)} \geq 3.52$, $p \leq 0.01$]. Further, participants demonstrated greater P3 amplitude over the center ROI during small problems ($m = 9.3 \pm 5.4$) relative to large problems ($m = 7.3 \pm 4.3$), [$t_{(39)} = 3.52$, $p < 0.01$]. Further, a fitness × problem size interaction was observed, [$F_{(1, 38)} = 6.33$, $p = 0.02$, $\eta^2 = 0.14$]. *Post-hoc* tests revealed that higher fit children ($m = 8.1 \pm 2.8$) had smaller

Table 3 | Amplitude and latency data for the P3 and N400 components for higher fit and lower fit children.

P3 amplitude (µv)	Higher fit	Lower fit
Small-correct-left	7.2 (3.7)	10.7 (5.1)
Small-incorrect-left	6.5 (4.5)	12.5 (4.9)
Large-correct-left	7.8 (4.3)	8.1 (5.5)
Large-incorrect-left	6.4 (4.4)	7.8 (5.2)
Small-correct-center	8.4 (4.2)	9.9 (6.4)
Small-incorrect-center	7.7 (4.0)	11.0 (7.0)
Large-correct-center	7.5 (4.5)	7.9 (4.3)
Large-incorrect-center	6.4 (4.4)	7.3 (4.8)
Small-correct-right	9.9 (4.2)	11.4 (7.0)
Small-incorrect-right	8.7 (4.1)	11.0 (7.0)
Large-correct-right	8.4 (5.1)	8.7 (5.9)
Large-incorrect-right	7.9 (5.0)	9.5 (6.5)
P3 latency (ms)		
Small-correct-left	371.5 (71.1)	362.4 (57.8)
Small-incorrect-left	365.9 (76.6)	365.5 (45.3)
Large-correct-left	368.1 (71.6)	394.0 (48.0)
Large-incorrect-left	381.7 (41.6)	389.3 (71.9)
Small-correct-center	353.9 (56.8)	370.0 (68.4)
Small-incorrect-center	405.2 (87.8)	379.6 (60.0)
Large-correct-center	408.4 (68.0)	401.2 (62.7)
Large-incorrect-center	418.0 (79.4)	411.0 (69.0)
Small-correct-right	342.1 (55.2)	356.3 (63.0)
Small-incorrect-right	376.4 (83.0)	355.8 (51.6)
Large-correct-right	361.7 (58.3)	374.5 (60.5)
Large-incorrect-right	361.8 (74.4)	389.0 (74.3)
N400 amplitude (µv)		
Small-correct-left	2.0 (4.6)	5.3 (4.1)
Small-incorrect-left	0.1 (5.2)	5.3 (4.2)
Large-correct-left	2.2 (4.9)	4.0 (4.3)
Large-incorrect-left	−0.2 (4.1)	3.7 (4.5)
Small-correct-right	2.5 (5.4)	5.4 (6.0)
Small-incorrect-right	0.4 (4.6)	0.3 (6.5)
Large-correct-right	2.6 (3.8)	5.2 (3.8)
Large-incorrect-right	0.7 (4.8)	3.5 (5.2)
N400 latency (ms)		
Small-correct-left	387.7 (58.5)	374.6 (28.4)
Small-incorrect-left	392.0 (79.7)	375.7 (30.8)
Large-correct-left	399.9 (77.9)	378.8 (30.9)
Large-incorrect-left	399.4 (77.8)	377.9 (34.7)
Small-correct-right	396.9 (81.2)	380.6 (27.7)
Small-incorrect-right	390.4 (69.0)	381.7 (20.1)
Large-correct-right	390.6 (63.7)	374.4 (25.8)
Large-incorrect-right	404.7 (70.3)	377.8 (33.2)

µv, microvolts; ms, milliseconds.

P3 amplitude relative to lower fit children ($m = 11.4 \pm 5.5$) during small problems, [$t_{(38)} = 2.36$], $p = 0.02$. Lastly, a fitness × correctness interaction indicated that lower fit ($m = 10.1 \pm 5.3$) relative to higher fit ($m = 7.3 \pm 2.7$) children exhibited larger P3 amplitude during incorrect problems, [$F_{(1, 38)} = 8.13$, $p = 0.002$, $\eta^2 = 0.17$]. However, *post-hoc* tests failed to decompose the interaction, [$t's_{(38)} \leq 2.16$, $p's \geq 0.04$].

P3 latency analyses revealed effects of problem size, [$F_{(1, 38)} = 10.50$, $p < 0.01$, $\eta^2 = 0.28$], indicating that all participants demonstrated longer P3 latency during large ($m = 388.2 \pm 46.7$) relative to small ($m = 367.0 \pm 43.0$) problems. Analysis further revealed an effect of correctness, [$F_{(1, 38)} = 3.96$, $p = 0.05$, $\eta^2 = 0.09$], indicating that participants had longer P3 latency during incorrect ($m = 383.2 \pm 48.7$) relative to correct ($m = 371.9 \pm 39.7$) problems. An effect of region was also observed, [$F_{(1, 38)} = 5.98$, $p < 0.01$, $\eta^2 = 0.14$], indicating that P3 latency was longest over the center ROI ($m = 393.4 \pm 52.7$), and shortest over the right ROI ($m = 347.6 \pm 50.8$), with the left ROI ($m = 374.8 \pm 48.8$) falling in-between, [$t_{(39)} = 4.00$, $p < 0.01$]. Lastly, a fitness $\times$ size $\times$ correctness $\times$ region interaction was observed, [$F_{(1, 38)} = 4.02$, $p = 0.02$, $\eta^2 = 0.1$], however, *post-hoc* tests failed to decompose the interaction, [$t's_{(38)} \leq 1.3$, $p's \geq 0.35$].

N400

N400 amplitude and latency data are presented in **Table 3**. Analysis revealed effects of fitness, [$F_{(1, 38)} = 6.40$, $p = 0.02$, $\eta^2 = 0.14$], and correctness, [$F_{(1, 38)} = 14.72$, $p < 0.01$, $\eta^2 = 0.28$], which were superseded by a fitness $\times$ correctness interaction, [$F_{(1, 38)} = 8.25$, $p < 0.01$, $\eta^2 = 0.18$]. *Post-hoc* testing revealed that higher fit ($m = 0.2 \pm 0.9$) relative to lower fit ($m = 4.4 \pm 1.0$) children had larger N400 amplitude during incorrect problems, [$t_{(38)} = 2.96$, $p < 0.01$]. Analysis based on difference waves (incorrect-correct) confirmed this finding, revealing an effect of fitness, [$F_{(1, 38)} = 8.25$, $p < 0.01$, $\eta^2 = 0.18$], indicating that higher fit children ($m = -2.1 \pm 2.0$) had greater N400 amplitude than lower fit children ($m = -0.3 \pm 1.8$) during incorrect solutions. No effects of fitness, problem size, ROI, or correctness were observed for N400 latency, [$F's_{(1, 38)} \leq 2.29$, $p's \geq 0.14$, $\eta's^2 \leq 0.02$].

DISCUSSION

The aim of the current study was to extend the literature-base in cardiorespiratory fitness and cognition by assessing strategic, behavioral, and electrophysiological indices of arithmetic cognition in preadolescent children. Consistent with *a priori* predictions, higher fit children reported using retrieval strategies more often for large problems compared to lower fit children; however, all children reported relying more on retrieval strategies for small relative to large problems, suggesting that fitness has a selective relation with specific aspects of arithmetic cognition. Alternatively, no fitness differences were observed for standardized achievement. During the verification task, fitness primarily modulated performance for large problems, but all children demonstrated behavioral modulations as a function of problem size and solution correctness. On the electrophysiological level, both early and late components were modulated by fitness and all participants demonstrated modulations of multiple ERP components as a function of problem size and solution correctness. Thus, these findings extend the current knowledge base of aerobic fitness-related benefits during neurocognitive development and add to a growing body of research detailing the development of arithmetic cognition.

STRATEGY

Higher fit children reported greater use of retrieval strategies than their lower fit counterparts during large problem performance, revealing fitness-related differences in strategic deployment as a function of problem size. Beyond fitness, all children reported more frequent retrieval for small relative to large problems. Differences in arithmetic strategy selection are believed to reflect the underlying functional integration of higher-order neurocognitive functions such as memory, visuo-spatial ability, and cognitive control (Grabner et al., 2007; Wu et al., 2009); functions that are known to develop across childhood (Holmes et al., 2009; Luna, 2009; Dumontheil and Klingberg, 2012) and which are positively influenced by fitness (Chaddock et al., 2011; Pontifex et al., 2011; Hillman et al., 2012; Monti et al., 2012). Accordingly, the current data provide evidence to suggest that fitness may positively influence strategy selection during arithmetic performance by benefiting the underlying cognitive constructs necessary for mature strategic implementation. To the best of our knowledge, these are the first data to demonstrate shifts in arithmetic strategy as a function of fitness, and raise interesting questions regarding possible differential neural underpinnings sub-serving strategic implementation between higher- and lower-fit children.

ACHIEVEMENT

Contrary to our predictions and in opposition to previous research (California Department of Education, 2001, 2005; Castelli et al., 2007; Wittberg et al., 2012), no differences in achievement were observed as a function of fitness level. While perplexing, this result may be due to the fact that the current sample was comprised of relatively high math achievers, whom demonstrated both above average IQ and SES; factors known to mediate mathematical achievement (White, 1982; Sirin, 2005). It is also possible that differences in the sensitivity and specificity between standardized achievement tests employed in current and past research, may in part, account for this discrepancy. Further research is necessary to clarify the relation between fitness and performance on standardized tests of mathematical achievement.

While no effects were observed with respect to fitness, all children did perform better on the math concepts, relative to math computation, subsection of the KTEA-2. Conceptual arithmetic knowledge is a prerequisite for inferential and adaptive arithmetic expertise (Hatano, 1988; Domahs and Delazer, 2005), providing a fundamental understanding of arithmetic operations and principals (Domahs and Delazer, 2005). Computational knowledge, while building on conceptual knowledge, also requires procedural guidance of algorithm execution known as routine expertise (Hatano, 1988), as well as the retrieval of declarative facts (Ashcraft, 1987; Siegler, 1988; Campbell, 1995), which arises from a synergy of conceptual and procedural mathematical knowledge (Domahs and Delazer, 2005). As such, it is not surprising that 9–10 year old children demonstrated superior performance for conceptual relative to computation achievement, as the latter naturally develops upon conceptual foundations.

ARITHMETIC VERIFICATION PERFORMANCE

Comparison of d' scores between fitness groups revealed greater performance during large problems for higher- relative

to lower-fit children. Furthermore, all children demonstrated decreased accuracy for large relative to small problems. Current explanations of the problem size effect attribute this phenomenon to differences in strategic deployment between large and small problems (Campbell and Xue, 2001; Zbrodoff and Logan, 2005), with less frequent and less efficient use of retrieval strategies for large relative to small problems. This results in greater interference between correct and incorrect solutions as problem sizes increase (Campbell and Xue, 2001; Campbell and Epp, 2004; Zbrodoff and Logan, 2005). As lower fit children reported relying on procedural strategies more frequently for large problems than their higher fit peers, lower fit children may have incurred a response criterion deficit, experiencing greater interference when attempting to detect correct and reject incorrect solutions. While novel to the arithmetic literature, differences in strategy implementation and interference control between higher- and lower-fit children is a common finding, with higher fit children regularly demonstrating more efficient and flexible strategy deployment, and superior interference control during experimental paradigms (Hillman et al., 2009; Pontifex et al., 2011; Voss et al., 2011; Chaddock et al., 2012). However, this is the first study to extend this finding to the domain of arithmetic. Thus, the beneficial influence of fitness on strategic deployment and interference control may confer neurocognitive benefits that translate across a variety of domains, including those necessary for arithmetic and academic success.

In addition, all children responded less accurately for incorrect relative to correct solutions, irrespective of problem size. Explanations for the split effect are less transpicuous than the problem size effect, as several plausible theories have been proposed (Campbell, 1987; Siegler, 1988; El Yagoubi et al., 2003; Duverne and Lemaire, 2005). Specifically, some researchers cite interference (Campbell, 1987), or frequency and strength of association between incorrect and correct solutions (Siegler, 1988), while others cite differences in verification strategy between correct and incorrect solutions (El Yagoubi et al., 2003; Duverne and Lemaire, 2005). Irrespective of cause, the current results provide information regarding the split effect during development, and more importantly, illustrate the interaction of the problem size and split effect (all children exhibited the poorest accuracy for large-incorrect problems). Accordingly, the current results provide an impetus for studying this interaction, particularly as the split and problem size effects, while well studied, are typically evaluated separately. Further evaluation of the combinatorial influence of the problem size and split effects will yield a finer understanding of arithmetic competency during development.

ERPs

Although no specific predictions were made relative to the early ERP components, several notable modulations as a function of fitness and task parameters occurred. First, while the P1 component is typically unevaluated in arithmetic verification paradigms, the current results suggest that fitness, solution correctness, and problem size may modulate P1 amplitude in children (see **Figures 2–4**). Specifically, although fitness significantly interacted with solution correctness, subsidiary analyses failed

to decompose into significant differences among the groups. However, the moderate effect sizes across ROIs ($0.68 > d > 0.30$) suggest significant effects may emerge in a larger sample (see **Figure 3**). Furthermore, children in the current study exhibited greater P1 amplitude during small relative to large solutions, and for incorrect relative to correct solutions. While P1 amplitude modulations as a function of solution size may be attributed to differing physical properties or spatial distributions of attention between small (e.g., 9) and large (e.g., 17; Mangun and Hillyard, 1991; Luck et al., 1994; Muluh et al., 2011) solutions, neither physical properties nor attentional distribution can account for amplitude modulations as a function of solution correctness (see **Figure 2**). As such, further research appears necessary to elucidate the meaning and theoretical implications of P1 amplitude modulations during arithmetic verification in relation to fitness and task parameters.
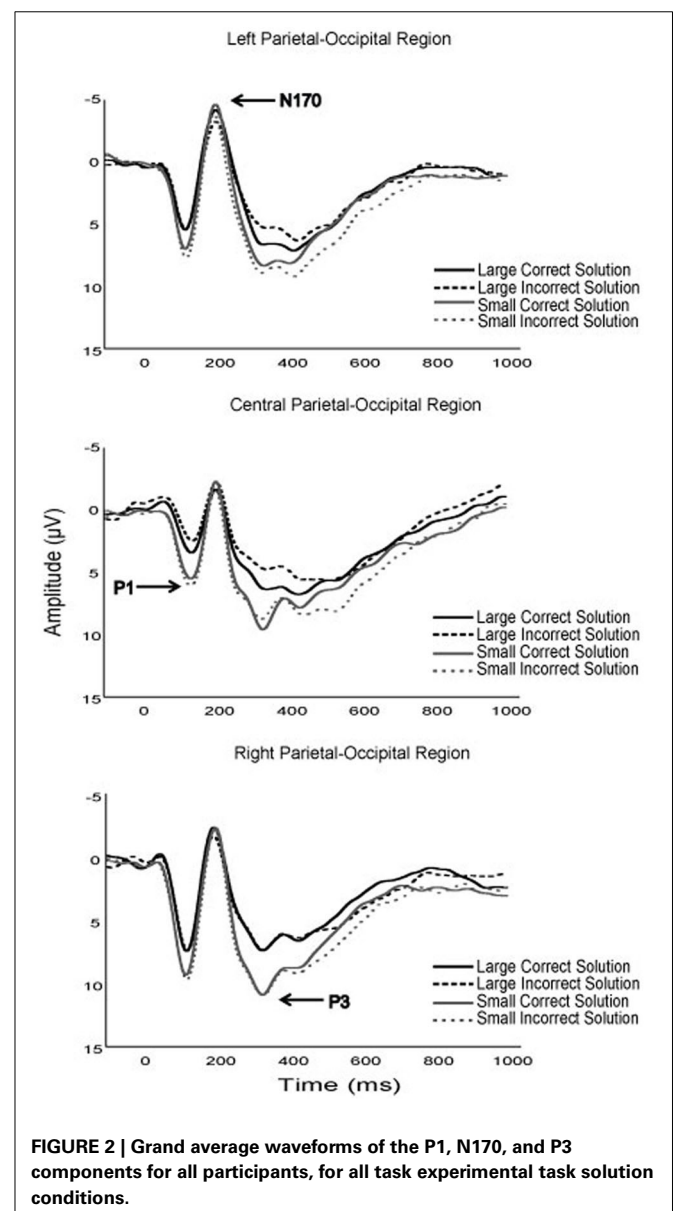

FIGURE 2 | Grand average waveforms of the P1, N170, and P3 components for all participants, for all task experimental task solution conditions.

FIGURE 3 | Grand average waveforms of the P1, N170, and P3 components for higher and lower fit participants, for correct and incorrect experimental task solutions.

FIGURE 4 | Grand average waveforms of the P1, N170 and P3 components for higher and lower fit participants, for large and small experimental task solutions.
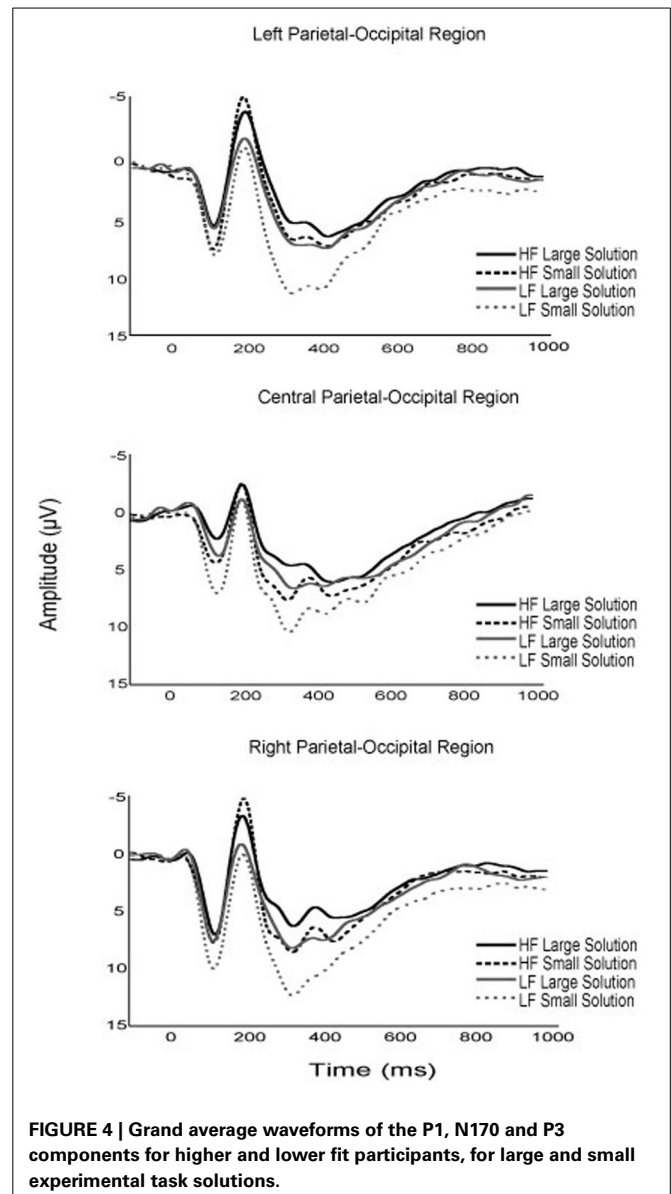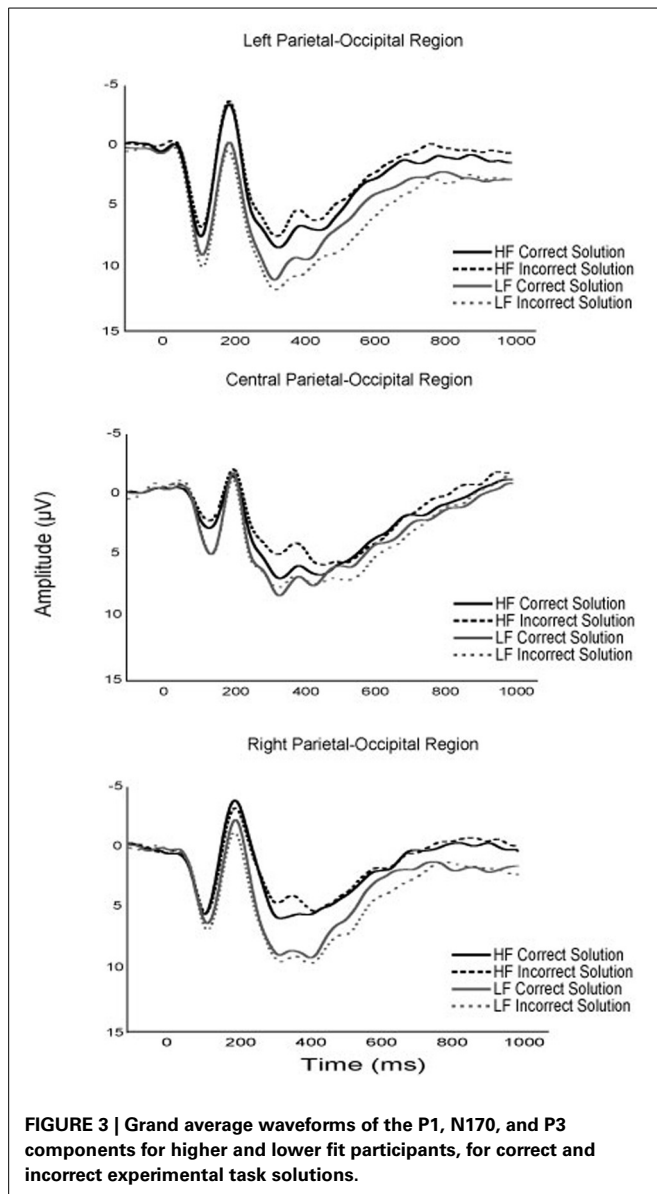
Secondly, higher fit children demonstrated greater N170 amplitude than their lower fit peers, and this group difference was found to interact with solution correctness, such that higher fit children demonstrated the greatest amplitude difference during incorrect solution processing (see **Figure 3**). The left lateralization of the N170 across participants observed herein links this component to the parietal-occipital N170 believed to reflect experience-dependent changes in visual expertise (Gauthier et al., 2003; Schlaggar and McCandliss, 2007; Maurer et al., 2008). Within the context of arithmetic verification, it has been suggested that the N170 reflects numeric symbol encoding (He et al., 2011). As such, the N170 observed during arithmetic verification may be an index of experience-dependent expertise in numeric symbol encoding. Fitness thus appears to benefit the neural resources responsible for numeric symbol encoding, with a disproportionate benefit for encoding incorrect solutions.
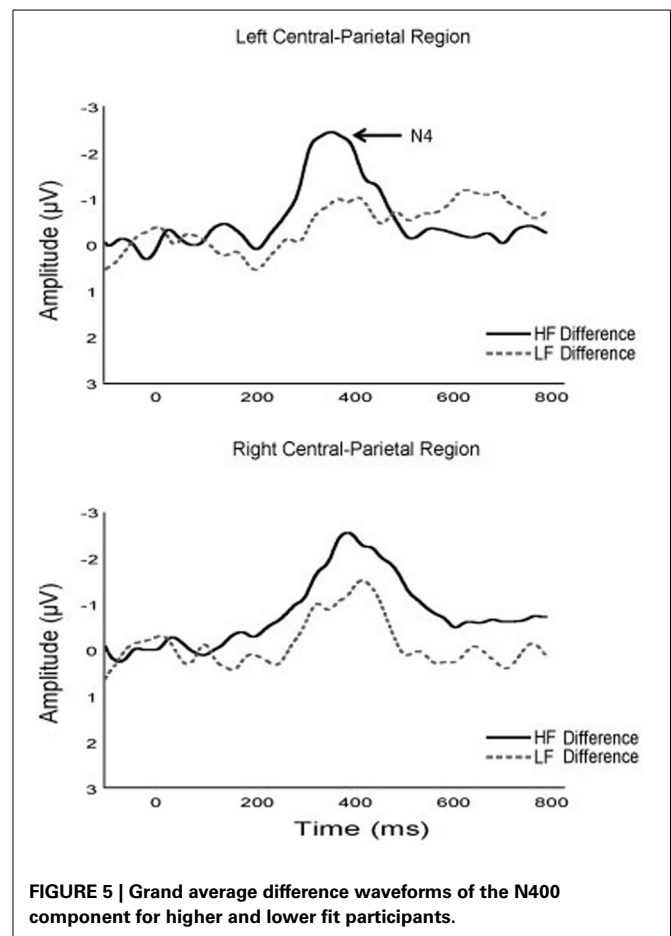
Post-hoc explanations of these data suggest that fitness may expedite the maturation of arithmetic expertise by facilitating differential numeric encoding of correct and incorrect solutions.

Both animal (van Praag et al., 1999; Cotman and Berchtold, 2002, 2007; van Praag, 2008) and human (Colcombe and Kramer, 2003; Kramer and Erickson, 2007; Chaddock et al., 2011; Erickson et al., 2011; Voss et al., 2011; Monti et al., 2012; Chaddock-Heyman et al., 2013) studies demonstrate the beneficial effects of cardiorespiratory fitness on experience-dependent changes in plasticity, connectivity, and integrity of a variety of cortical and subcortical areas. Furthermore, neural structures and networks critical for arithmetic cognition, such as the hippocampus (Rivera et al., 2005; Cho et al., 2011; De Smedt et al., 2011), prefrontal and posterior parietal cortices (Dehaene et al., 2003; Nieder and Dehaene, 2009; Cho et al., 2011; De Smedt et al., 2011), and the fronto-parietal network (Dehaene et al., 2003; Nieder and

Dehaene, 2009), show disproportionate fitness-related benefits (Colcombe and Kramer, 2003; Colcombe et al., 2004a,b; Hillman et al., 2008; Erickson and Kramer, 2009; Chaddock et al., 2011; Erickson et al., 2011; Voss et al., 2011). Therefore, fitness may facilitate experience-dependent changes in the neural architecture sub-serving numeric symbol encoding, resulting in the functional electrophysiological alterations currently observed. Future multimodal research will be well positioned to further elucidate the neural specificity of this relation during arithmetic performance.

With respect to later ERP components, lower- relative to higher-fit children exhibited greater P3 amplitude during small problem solutions, with the greatest difference occurring for small-incorrect solutions (see **Figures 3**, **4**). While all participants exhibited greater P3 amplitude for small relative to large problems, the current fitness finding suggests that small problems, required greater attentional resources for lower- relative to higher-fit children. Stated differently, higher fit children were able to maintain equivalent performance for small problems, irrespective of solution correctness, while up-regulating fewer attentional resources relative to their lower fit peers. The current results add to those of Wu and Hillman (2013), and provide further evidence that pediatric fitness is associated with more flexible attentional resource allocation in relation to task demands. Further evidence is provided by research examining pediatric fitness and brain function on the hemodynamic level (Chaddock et al., 2012; Chaddock-Heyman et al., 2013), which demonstrate that higher fit children exhibit more efficient neural resource allocation in relation to task demands during a task requiring attentional inhibition and interference control. Given, the variety of tasks (i.e., attentional blink, arithmetic verification, flanker) and multimodal (ERP, fMRI) convergence, it appears that higher fit children may derive a generalizable benefit across tasks through optimizing attentional resource allocation in relation to task demands.

In addition to P3 amplitude modulations, higher fit children exhibited significantly greater N400 amplitude to incorrect solutions relative to their lower fit counterparts; a finding further confirmed by difference wave analysis (see **Figure 5**). Accordingly, fitness appears to influence semantic memory processing during arithmetic verification. Further, tertiary analysis revealed that d' scores were positively correlated with N400 amplitude, suggesting that fitness may facilitate the detection of correct solutions and rejection of incorrect solutions via differential activation of semantic memory networks. Indeed, the only other study to evaluate the underlying neurocognitive processes giving rise to greater achievement scores in higher fit children observed a similar finding within the domain of linguistic performance (Scudder et al., 2014). In this study, behavioral and electrophysiological function in higher- and lower-fit children was observed as they read sentences that were either semantically or syntactically congruent (correct) or incongruent (incorrect). In addition to exhibiting shorter RT, higher- relative to lower-fit children exhibited greater N400 amplitude and shorter N400 latency; suggesting that cardiorespiratory fitness during development facilitates the extraction of semantic information during sentence reading. Thus, the current results both compliment and extend the results of Scudder et al. (2014), which together suggest that fitness

FIGURE 5 | Grand average difference waveforms of the N400 component for higher and lower fit participants.

positively relates to semantic processing during academic-based tasks. The N400 therefore appears to be a convergent electrophysiological mechanism supporting fitness-related benefits observed across academic domains.

LIMITATIONS AND CONCLUSION

While the comprehensive nature of the current study yields valuable information regarding the relation of cardiorespiratory fitness to aspects of arithmetic cognition, it is not without limitations. First, the study design was cross-sectional in nature and it is always possible that some unmeasured variable may have influenced the current results. However, demographic variables such as age, IQ, SES and pubertal timing did not differ between groups and were relatively homogenous between participants. In addition, the relatively small sample size may limit the interpretable power of the current results. Future longitudinal studies with greater sample size will help determine the robustness of the observed effects. Lastly, the current sample was relatively high performing in terms of IQ and academic achievement, potentially limiting the generalizability of the current results.

Irrespective of limitations, the findings observed herein add important information to the fitness-cognition literature by revealing that the beneficial effects of fitness extend on the behavioral and neural levels to the domain of arithmetic cognition. The current results provide further incentive for promoting physical

activity and fitness in youth, while engendering further inquiry into the relation of fitness and scholastic development. By further detailing strategic, behavioral, and electrophysiological indices of arithmetic cognition during development, the current results also call for a more refined examination of arithmetic development through the evaluation of early ERP components during arithmetic verification as well as the interaction of size and split effects. In summary, the current results add important information to the exercise and arithmetic cognition literatures, illustrating the importance of a physically active lifestyle as well as comprehensive experimental designs when evaluating scholastic development. Lastly, the current results further emphasize the importance of cardiorespiratory fitness during childhood not only for cardiovascular health, but also for neurocognitive and scholastic development.

REFERENCES

American College of Sports Medicine. (2010). *ACSM's Guidelines for Exercise Testing and Prescription (7th Edn.)*. New York, NY: Lippincott Williams and Wilkins.

Ashcraft, M. H. (1982). The development of mental arithmetic: achronometric approach. *Dev. Rev.* 2, 213–236. doi: 10.1016/0273-2297(82)90012-0

Ashcraft, M. H., and Fierman, B. A. (1982). Mental addition in third, fourth, and sixth graders. *J. Exp. Child Psychol.* 33, 216–234. doi: 10.1016/0022-0965(82)90017-0

Ashcraft, M. H. (1987). "Children's knowledge of simple arithmetic: a developmental model and simulation," in *Formal Methods in Developmental Research*, eds R. Brainerd and J. Bisanz (New York, NY: Springer-Verlag), 302–338.

Bar-Or, O. (1983). *Pediatric Sports Medicine for the Practitioner: From Physiologic Principles to Clinical Applications*. New York, NY: Springer-Verlag. doi: 10.1007/978-1-4612-5593-2

Birnbaum, A. S., Lytle, L. A., Murray, D. M., Story, M., Perry, C. L., and Boutelle, K. N. (2002). Survey development for assessing correlates of young adolescents' eating. *Am. J. Health Behav.* 26, 284–295. doi: 10.5993/AJHB.26.4.5

Buck, S. M., Hillman, C. H., and Castelli, D. M. (2008). The relation of aerobic ?tness to stroop task performance in preadolescent children. *Med. Sci. Sports Exerc.* 40, 166–172. doi: 10.1249/mss.0b013e318159b035

California Department of Education. (2001). *California Physical Fitness Test: Report to the Governor and Legislature*. Sacramento, CA: California Department of Education Standards and Assessment Division.

California Department of Education. (2005). *California Physical Fitness Test: Report to the Governor and Legislature*. Sacramento, CA: California Department of Education Standards and Assessment Division.

Campbell, J. I., and Epp, L. J. (2004). An encoding-complex approach to numerical cognition in Chinese-English bilinguals. *Can. J. Exp. Psychol.* 58, 229. doi: 10.1037/h0087447

Campbell, J. I. D., and Fugelsang, J. (2001). Strategy choice for arithmetic verification: effects of numerical surface form. *Cognition* 80, 21–30. doi: 10.1016/S0010-0277(01)00115-9

Campbell, J. I. D., and Xue, Q. L. (2001). Cognitive arithmetic across cultures. *J. Exp. Psychol. Gen.* 130, 299–315. doi: 10.1037/0096-3445.130.2.299

Campbell, J. I. (1995). Mechanisms of simple addition and multiplication: a modified network interference theory and simulation. *Math. Cogn.* 1, 121–165.

Campbell, J. I. D. (1987). Network interference and mental multiplication. *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 109–123. doi: 10.1037/0278-7393.13.1.109

Castelli, D. M., Hillman, C. H., Buck, S. M., and Erwin, H. E. (2007). Physical fitness and academic achievement in third-and fifth-grade students. *J. Sport Exerc. Psychol.* 29:239.

Chaddock-Heyman, L., Erickson, K. I., Voss, M. W., Powers, J. P., Knecht, A. M., Pontifex, M. B., et al. (2013). White matter microstructure is associated with cognitive control in children. *Biol. Psychol.* 94, 109–115. doi: 10.1016/j.biopsycho.2013.05.008

Chaddock, L., Erickson, K. I., Prakash, R. S., Voss, M. W., VanPatter, M., Pontifex, M. B., et al. (2012). A functional MRI investigation of the association between childhood aerobic fitness and neurocognitive control. *Biol. Psychol.* 89, 260–268. doi: 10.1016/j.biopsycho.2011.10.017

Chaddock, L., Hillman, C. H., Buck, S. M., and Cohen, N. J. (2011). Aerobic fitness and executive control of relational memory in preadolescent children. *Med. Sci. Sports Exerc.* 43, 344–349. doi: 10.1249/MSS.0b013e318 1e9af48

Chatrian, G. E., Lettich, E., and Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *Am. J. Electroencephalogr. Clin. Neurophysiol.* 25, 83–92.

Chen, Y., Campbell, J. I., and Liu, C. (2013). The N3 is sensitive to odd–even congruency information in arithmetic fact retrieval. *Exerc. Brain Res.* 255, 603–611. doi: 10.1007/s00221-013-3404-9

Cho, S., Ryali, S., Geary, D. C., and Menon, V. (2011). How does a child solve 7+ 8? Decoding brain activity patterns associated with counting and retrieval strategies. *Dev. Sci.* 14, 989–1001. doi: 10.1111/j.1467-7687.2011.01055.x

Chomitz, V. R., Slining, M. M., McGowan, R. J., Mitchell, S. E., Dawson, G. F., and Hacker, K. A. (2009). Is there a relationship between physical fitness and academic achievement? Positive results from public school children in the northeastern United States. *J. Sch. Health* 79, 30–37. doi: 10.1111/j.1746-1561.2008.00371.x

Colcombe, S. J., Kramer, A. F., Erickson, K. I., Scalf, P., McAuley, E., Cohen, N. J., et al. (2004a). Cardiovascular fitness, cortical plasticity, and aging. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3316–3321. doi: 10.1073/pnas.0400266101

Colcombe, S. J., Kramer, A. F., McAuley, E., Erickson, K. I., and Scalf, P. (2004b). Neurocognitive aging and cardiovascular fitness. *J. Mol. Neurosci.* 24, 9–14. doi: 10.1385/JMN:24:1:009

Colcombe, S., and Kramer, A. F. (2003). Fitness effects on the cognitive function of older adults A meta-analytic study. *Psychol. Sci.* 14, 125–130. doi: 10.1111/1467-9280.t01-1-01430

Cotman, C. W., and Berchtold, N. C. (2002). Exercise: a behavioral intervention to enhance brain health and plasticity. *Trends Neurosci.* 25, 295–301. doi: 10.1016/S0166-2236(02)02143-4

Cotman, C. W., and Berchtold, N. C. (2007). Physical activity and the maintenance of cognition: learning from animal models. *Alzheimers Dement.* 3, S30–S37. doi: 10.1016/j.jalz.2007.01.013

Cottrell, L. A., Northrup, K., and Wittberg, R. (2007). The extended relationship between child cardiovascular risks and academic performance measures. *Obesity* 15, 3170–3177. doi: 10.1038/oby.2007.377

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239

Dehaene, S. (1996): The organization of brain activations in number comparison: event-related potentials and the additive-factors method. *J. Cogn. Neurosci.* 8, 47–68. doi: 10.1162/jocn.1996.8.1.47

De Smedt, B., Holloway, I. D., and Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetical fluency. *Neuroimage* 57, 771–781. doi: 10.1016/j.neuroimage.2010.12.037

De Smedt, B., Taylor, J., Archibald, L., and Ansari, D. (2010). How is phonological processing related to individual differences in children's arithmetic skills? *Dev. Sci.* 13, 508–520. doi: 10.1111/j.1467-7687.2009.00897.x

Domahs, F., Domahs, U., Schlesewsky, M., Ratinckx, E., Verguts, T., Willmes, K., et al. (2007). Neighborhood consistency in mental arithmetic: Behavioral and ERP evidence. *Behav. Brain Funct.* 3, 66. doi: 10.1186/1744-9081-3-66

Domahs, F., and Delazer, M. (2005). Some assumptions and facts about arithmetic facts. *Psychol. Sci.* 47, 96–111.

Dumontheil, I., and Klingberg, T. (2012). Brain activity during a visuospatial working memory task predicts arithmetical performance 2 years later. *Cereb. Cortex* 22, 1078–1085. doi: 10.1093/cercor/bhr175

DuPaul, G. T., Power, T. J., Anastopoulos, A. D., and Reid, R. (1998). *ADHD Rating Scale, IV: Checklists, Norms, and Clinical Interpretation*. New York, NY: Guilford Press.

Duverne, S., and Lemaire, P. (2005). Arithmetic split effects re?ect strategy selection: an adult age comparative study in addition comparison and veri?cation tasks. *Can. J. Exp. Psychol.* 59, 262–278. doi: 10.1037/h0087479

El Yagoubi, R., Lemaire, P., and Besson, M. (2003). Different brain mechanisms mediate two strategies in arithmetic: evidence from event related brain potentials. *Neuropsychologia* 41, 855–862. doi: 10.1016/S0028-3932(02)00180-X

El Yagoubi, R., Lemaire, P., and Besson, M. (2005). Effects of aging on arithmetic problem-solving: an event-related brain potential study. *J. Cogn. Neurosci.* 17, 37–50. doi: 10.1162/0898929052880084

Erickson, K. I., and Kramer, A. F. (2009). Aerobic exercise effects on cognitive and neural plasticity in older adults. *Br. J. Sports Med.* 43, 22–24. doi: 10.1136/bjsm.2008.052498

Erickson, K. I., Voss, M. W., Prakash, R. S., Basak, C., Szabo, A., Chaddock, L., et al. (2011). Exercise training increases size of hippocampus and improves memory. *Proc. Nat. Acad. Sci. U.S.A.* 108, 3017–3022. doi: 10.1073/pnas.1015950108

Federmeier, K. D., and Laszlo, S. (2009). "Time for meaning: electrophysiology provides insights into the dynamics of representation and processing in semantic memory," in *The Psychology of Learning and Motivation*, ed B. H. Ross (San Diego, CA: Elsevier), 1–44.

Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307–314. doi: 10.1016/j.tics.2004.05.002

Freedson, P. S., and Goodman, T. L. (1993). "Measurement of oxygen consumption," in *Pediatric Laboratory Exercise Testing: Clinical Guidelines*, ed T. W. Rowland (Champaign, IL: Human Kinetics), 91–113.

Galfano, G., Mazza, V., Angrilli, A., and Umilta, C. (2004). Electrophysiological correlates of stimulus-driven multiplication facts retrieval. *Neuropsychologia* 42, 1370–1382 doi: 10.1016/j.neuropsychologia.2004.02.010

Gauthier, I., Curran, T., Curby, K. M., and Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nat. Neurosci.* 6, 428–432. doi: 10.1038/nn1029

Geary, D. C. (2010). Mathematical disabilities: reflections on cognitive, neuropsychological, and genetic components. *Learn. Individ. Differ.* 20, 130–133. doi: 10.1016/j.lindif.2009.10.008

Geary, D. C., Hoard, M. K., Byrd-Craven, J., and DeSoto, M. C. (2004). Strategy choices in simple and complex addition: contributions of working memory and counting knowledge for children with mathematical disability. *J. Exp. Child Psychol.* 88, 121–151. doi: 10.1016/j.jecp.2004.03.002

Geary, D., Widaman, K., Little, T., and Cormier, P. (1987). Cognitive addition: comparison of learning disabled and academically normal elementary school children. *Cogn. Dev.* 2, 249–269. doi: 10.1016/S0885-2014(87)90075-X

Grabner, R. H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., and Neuper, C. (2007). Individual differences in mathematical competence predicts parietal brain activation during mental calculation. *Neuroimage* 38, 346–356. doi: 10.1016/j.neuroimage.2007.07.041

Groen, G., and Parkman, J. (1972). Chronometric analysis of simple addition. *Psychol. Rev.* 79, 329–343. doi: 10.1037/h0032950

Hatano, G. (1988). "Social and motivational bases for mathematical understanding," in *Children's Mathematics*, eds G. B. Saxe and M. Gearhart (San Francisco, Ca: Jossey-Bass), 35–50.

He, W. Q., Luo, W. B., He, H. M., Chen, X., and Zhang, D. J. (2011). N170 effects during exact and approximate calculation tasks: an ERP study. *Neuroreport* 22, 437–441. doi: 10.1097/WNR.0b013e32834702c1

Heine, A., Wißmann, J., Tamm, S., De Smedt, B., Schneider, M., Stern, E., et al. (2012). An electrophysiological investigation of non-symbolic magnitude processing: numerical distance effects in children with and without mathematical learning disabilities. *Cortex* 49, 3238–3246. doi: 10.1016/j.cortex.2012.11.009

Hillman, C. H., Buck, S. M., Themanson, J. R., Pontifex, M. B., and Castelli, D. M. (2009). Aerobic ?tness and cognitive development: event-related brain potential and task performance indices of executive control in preadolescent children. *Dev. Psychol.* 45, 114–129. doi: 10.1037/a0014437

Hillman, C. H., Castelli, D. M., and Buck, S. M. (2005). Aerobic ?tness and neurocognitive function in healthy preadolescent children. *Med. Sci. Sports Exerc.* 37, 1967–1974. doi: 10.1249/01.mss.0000176680.79702.ce

Hillman, C. H., Erickson, K. I., and Kramer, A. F. (2008). Be smart, exercise your heart: exercise effects on brain and cognition. *Nat. Rev. Neurosci.* 9, 58–65. doi: 10.1038/nrn2298

Hillman, C. H., Kamijo, K., and Pontifex, M. B. (2012). "The relation of ERP indices of exercise to brain health and cognition," in *Functional Neuroimaging in Exercise and Sport Sciences*, eds H. Boecker, C. H. Hillman, L. Scheef, and H. K. Strüder (New York, NY: Springer), 419–446. doi: 10.1007/978-1-4614-3293-7_18

Hillman, C. H., Motl, R. W., Pontifex, M. B., Posthuma, D., Stubbe, J. H., Boomsma, D. I., et al. (2006). Physical activity and cognitive function in a cross-section of younger and older community-dwelling individuals. *Health Psychol.* 25, 678–687. doi: 10.1037/0278-6133.25.6.678

Holmes, J., Gathercole, S. E., and Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Dev. Sci.* 12, F9–F15 doi: 10.1111/j.1467-7687.2009.00848.x

Hyde, D. C., and Spelke, E. S. (2009). All numbers are not equal: an electrophysiological investigation of small and large number representations. *J. Cogn. Neurosci.* 21, 1039–1053. doi: 10.1162/jocn.2009.21090

Hyde, D. C., and Spelke, E. S. (2012). Spatiotemporal dynamics of processing nonsymbolic number: an event–related potential source localization study. *Hum. Brain Mapp.* 33, 2189–2203. doi: 10.1002/hbm.21352

Imbo, I., and Vandierendonck, A. (2008). Effects of problem size, operation, and working-memory span on simple-arithmetic strategies: differences between children and adults? *Psychol. Res.* 72, 331–346. doi: 10.1007/s00426-007-0112-8

Institute of Medicine of the National Academies. (2013). *Educating the Student Body: Taking Physical Activity and Physical Education to School*. Washington, DC: The National Academies Press.

Jasinski, E. C., and Coch, D. (2012). ERPs across arithmetic operations in a delayed answer verification task. *Psychophysiology* 49, 943–958. doi: 10.1111/j.1469-8986.2012.01378.x

Jost, K., Hennighausen, E., and Rosler, F. (2004). Comparing arithmetic and semantic fact retrieval: effects of problem size and sentence constraint on event-related brain potentials. *Psychophysiology* 41, 46–59. doi: 10.1111/1469-8986.00119_41_1

Kaufman, A. S., and Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test*. Circle Pines, MN: American Guidance Service.

Kramer, A. F., and Erickson, K. I. (2007). Capitalizing on cortical plasticity: influence of physical activity on cognition and brain function. *Trends Cogn. Sci.* 11, 342–348. doi: 10.1016/j.tics.2007.06.009

Kramer, A. F., Erickson, K. I., and Colcombe, S. J. (2006). Exercise, cognition, and the aging brain. *J. Appl. Phys.* 101, 1237–1242. doi: 10.1152/japplphysiol.00500.2006

Kutas M., and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470 doi: 10.1016/S1364-6613(00)01560-6

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Ann. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., and Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: psychophysical and electrophysiological evidence for early selection. *J. Exp. Psychol.* 20, 887.

Luna, B. (2009). Developmental changes in cognitive control through adolescence. *Adv. Child Dev. Behav.* 37, 233–278. doi: 10.1016/S0065-2407(09)03706-9

Mangun, G. R., and Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *J. Exp. Psychol.* 17, 1057.

Maurer, U., Zevin, J. D., and McCandliss, B. D. (2008). Left-lateralized N170 effects of visual expertise in reading: evidence from Japanese syllabic and logographic scripts. *J. Cogn. Neurosci.* 20, 1878–1891. doi: 10.1162/jocn.2008.20125

Menon, V. (2010). Developmental cognitive neuroscience of arithmetic: implications for learning and education. *ZDM* 42, 515–525. doi: 10.1007/s11858-010-0242-0

Monti, J. M., Hillman, C. H., and Cohen, N. J. (2012). Aerobic fitness enhances relational memory in children: the FIT Kids randomized control trial. *Hippocampus* 22, 1876–1882. doi: 10.1002/hipo.22023

Moore, R. D., Wu, C. T., Pontifex, M. B., O'Leary, K. C., Scudder, M. R., Raine, L. B., et al. (2013). Aerobic fitness and intra-individual variability of neurocognition in preadolescent children. *Brain Cogn.* 82, 43–57. doi: 10.1016/j.bandc.2013.02.006

Muluh, E. T. (2011). A review of event-related potential (ERP) components employed in mental arithmetic processing studies. *J. Neurol. Neurophysiol.* S6, 1–7. doi: 10.4172/2155-9562.S6-001

Muluh, E. T., Vaughan, C. L., and John, L. R. (2011). High resolution event-related potentials analysis of the arithmetic-operation effect in mental arithmetic. *Clin. Neurophysiol.* 122, 518–529. doi: 10.1016/j.clinph.2010.08.008

Nieder, A., and Dehaene, S. (2009). Representation of number in the brain. *Annu. Rev. Neurosci.* 32, 185–208. doi: 10.1146/annurev.neuro.051508.135550

Niedeggen, M., and Rosler, F. (1999). N400 effects re?ect activation spread during retrieval of arithmetic facts. *Psychol. Sci.* 10, 271–276. doi: 10.1111/1467-9280.00149

Niedeggen, M., Rosler, F., and Jost, K. (1999). Processing of incongruous mental calculation problems: evidence for an arithmetic N400 effect. *Psychophysiology* 36, 307–324. doi: 10.1017/S0048577299980149

Núñez-Peña, M. I., Cortiñas, M., and Escera, C. (2006). Problem size effect and processing strategies in mental arithmetic. *Neuroreport* 17, 357–360. doi: 10.1097/01.wnr.0000203622.24953.c2

Núñez-Peña, M. I., Gracia-Bafalluy, M., and Tubau, E. (2011). Individual differences in arithmetic skill reflected in event-related brain potentials. *Int. J. Psychophysiol.* 80, 143–149. doi: 10.1016/j.ijpsycho.2011.02.017

Núñez-Peña, M. I., and Honrubia-Serrano, M. L. (2004). P600 related to rule violation in an arithmetic task. *Cogn. Brain Res.* 18, 130–141. doi: 10.1016/j.cogbrainres.2003.09.010

Núñez-Peña, M. I., Honrubia-Serrano, M. L., and Escera, C. (2004). Problem size effect in additions and subtractions: an event-related potential study. *Neurosci. Lett.* 373, 21–25. doi: 10.1016/j.neulet.2004.09.053

Núñez-Peña, M. I., and Suárez-Pellicioni, M. (2012). Processing false solutions in additions: differences between high-and lower-skilled arithmetic problem-solvers. *Exp. Brain Res.* 218, 655–663. doi: 10.1007/s00221-012-3058-z

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Palomares, M., Smith, P. R., Pitts, C. H., and Carter, B. M. (2011). The effect of viewing eccentricity on enumeration. *PLoS ONE* 6:e20779. doi: 10.1371/journal.pone.0020779

Pontifex, M. B., Hillman, C. H., and Polich, J. (2009). Age, physical fitness, and attention: P3a and P3b. *Psychophysiology* 46, 379–387. doi: 10.1111/j.1469-8986.2008.00782.x

Pontifex, M. B., Raine, L. B., Johnson, C. R., Chaddock, L., Voss, M. W., Cohen, N. J., et al. (2011). Cardiorespiratory ?tness and the ?exible modulation of cognitive control in preadolescent children. *J. Cogn. Neurosci.* 23, 1332–1345. doi: 10.1162/jocn.2010.21528

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

Rips, L. J., Bloomfield, A., and Asmuth, J. (2008). From numerical concepts to concepts of number. *Behav. Brain Sci.* 31, 623–687. doi: 10.1017/S0140525X08005566

Rivera, S. M., Reiss, A. L., Eckert, M. A., and Menon, V. (2005). Developmental changes in mental arithmetic: evidence for increased functional specialization in the left inferior parietal cortex. *Cereb. Cortex* 15, 1779–1790. doi: 10.1093/cercor/bhi055

Roussel, J. L., Fayol, M., and Barrouillet, P. (2002). Procedural vs direct retrieval strategies in arithmetic: A comparison between additive and multiplicative problem solving. *Eur. J. Cogn. Psychol.* 14, 61–104. doi: 10.1080/09541440042000115

Schlaggar, B. L., and McCandliss, B. D. (2007). Development of neural systems for reading. *Annu. Rev. Neurosci.* 30, 475–503. doi: 10.1146/annurev.neuro.28.061604.135645

Scudder, M. R., Federmeier, K. D., Raine, L. R., Direito, A., Boyd, J., and Hillman, C. H. (2014). The association between aerobic fitness and language processing in children: implications for academic achievement. *Brain Cogn.* 87C, 140–152. doi: 10.1016/j.bandc.2014.03.016

Szűcs, D., and Goswami, U. (2007). Educational neuroscience: defining a new discipline for the study of mental representations. *Mind Brain Educ.* 1, 114–127. doi: 10.1111/j.1751-228X.2007.00012.x

Shvartz, E., and Reibold, R. C. (1990). Aerobic fitness norms for males and females aged 6 to 75 years: a review. *Aviat. Space Environ. Med.* 61, 3–11.

Sibley, B. A., and Etnier, J. L. (2003). The relationship between physical activity and cognition in children: a meta-analysis. *Pediatr. Exerc. Sci.* 15, 243–256.

Siegler, R. (1988). Individual differences in strategy choices: good students, not-so good students, and perfectionists. *Child Dev.* 59, 833–851. doi: 10.2307/1130252

Siegler, R. S. (1987). The perils of averaging data over strategies: an example from children's addition. *J. Exp. Psychol. Gen.* 116, 250–264. doi: 10.1037/0096-3445.116.3.250

Siegler, R. S. (1986). "Unities across domains in children's strategy choices," in *Cognitive Process in Mathematics*, eds J. Sloboda and D. Rogers (Oxford: Oxford University Press), 81–106.

Siegler, R. S., and Shrager, J. (1984). "Strategy choice in addition and subtraction: how do children know what to do?," in *Origins of Cognitive Skills*, ed C. Sophian (Hillsdale, NJ: Erlbaum), 229–293.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Rev. Educ. Res.* 75, 417–453. doi: 10.3102/00346543075003417

Smith, P. J., Blumenthal, J. A., Hoffman, B. M., Cooper, H., Strauman, T. A., and Welshbohmer, K. (2010). Aerobic exercise and neurocognitive performance: a meta-analytic review of randomized controlled trials. *Psychosom. Med.* 72, 239–252. doi: 10.1097/PSY.0b013e3181d14633

Taylor, S. J., Whincup, P. H., Hindmarsh, P. C., Lampe, F., Odoki, K., and Cook, D. G. (2001). Performance of a new pubertal self-assessment questionnaire: a preliminary study. *Paediatr. Perinat. Epidemiol.* 15, 88–94. doi: 10.1046/j.1365-3016.2001.00317.x

Thomas, S., Reading, J., and Shephard, R. J. (1992). Revision of the physical activity readiness questionnaire (PAR-Q). *Can. J. Sport Sci.* 17, 338–345.

Tomporowski, P. D. (2003). Effects of acute bouts of exercise on cognition. *Acta Psychol.* 112, 297–324. doi: 10.1016/S0001-6918(02)00134-8

Utter, A. C., Roberson, R. J., Nieman, D. C., and Kang, J. (2002). Children's OMNI scale of perceived exertion: walking/running evaluation. *Med. Sci. Sports Exerc.* 34, 139–144. doi: 10.1097/00005768-200201000-00021

van Praag, H. (2008). Neurogenesis and exercise: past and future directions. *Neuromolecular Med.* 10, 128–140. doi: 10.1007/s12017-008-8028-z

van Praag, H., Kempermann, G., and Gage, F. H. (1999). Running increases cell proliferation and neurogenesis in the adult mouse dentate gyrus. *Nat. Neurosci.* 2, 266–270. doi: 10.1038/6368

Voss, M. S., Chaddock, L., Kim, J. S., VanPatter, M., Pontifex, M. B., Raine, L. B., et al. (2011). Aerobic fitness is associated with greater efficiency of the network underlying cognitive control in preadolescent children. *Neuroscience* 199, 166–176. doi: 10.1016/j.neuroscience.2011.10.009

White, K. (1982). The relation between socioeconomic status and academic achievement. *Psychol. Bull.* 91, 461–481. doi: 10.1037/0033-2909.91.3.461

Wu, C.-T., and Hillman, C. H. (2013). Aerobic fitness and the attentional blink in preadolescent children. *Neuropsychology* 27, 642–653. doi: 10.1037/a0034025

Wu, S. S., Chang, T. T., Majid, A., Caspers, S., Eickhoff, S. B., and Menon, V. (2009). Functional heterogeneity of inferiorparietal cortex during mathematical cognition assessed with cytoarchitectonic probability maps. *Cereb. Cortex* 19, 2930–2945. doi: 10.1093/cercor/bhp063

Wittberg, R. A., Northrup, K. L., and Cottrell, L. A. (2012). Children's aerobic fitness and academic achievement: a longitudinal examination of students during their fifth and seventh grade years. *Am. J. Public Health* 102, 2303–2307. doi: 10.2105/AJPH.2011.300515

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, 1–17. doi: 10.1080/09297040600760465

Zbrodoff, N. J., and Logan, G. D. (2005). "What everyone finds: the problem-size effect," in *The Handbook of Mathematical Cognition*, ed J. I. D. Campbell (Hove: Psychology Press), 331–345.