Brain and Cognition 87 (2014) 140–152

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

The association between aerobic fitness and language processing in children: Implications for academic achievement

Mark R. Scudder, Kara D. Federmeier, Lauren B. Raine, Artur Direito [1], Jeremy K. Boyd [2], Charles H. Hillman *

University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

ARTICLE INFO

Article history:
Accepted 24 March 2014

Keywords:
ERP
N400
Semantic processing
P600
Syntactic processing

ABSTRACT

Event-related brain potentials (ERPs) have been instrumental for discerning the relationship between children's aerobic fitness and aspects of cognition, yet language processing remains unexplored. ERPs linked to the processing of semantic information (the N400) and the analysis of language structure (the P600) were recorded from higher and lower aerobically fit children as they read normal sentences and those containing semantic or syntactic violations. Results revealed that higher fit children exhibited greater N400 amplitude and shorter latency across all sentence types, and a larger P600 effect for syntactic violations. Such findings suggest that higher fitness may be associated with a richer network of words and their meanings, and a greater ability to detect and/or repair syntactic errors. The current findings extend previous ERP research explicating the cognitive benefits associated with greater aerobic fitness in children and may have important implications for learning and academic performance.

1. Introduction

As much as a century ago, scientific evidence demonstrated higher rates of degenerative diseases in individuals with sedentary occupations (Paffenbarger, Blair, & Lee, 2001). Despite the refinement of this knowledge over subsequent decades, changes in technology and lifestyle have led to continued or even increased prevalence of sedentary behaviors in individuals of all ages (Vayman & Gomez-Pinilla, 2006), with arguably the most negative outcomes accruing to children and adolescents. Since 1980, the number of overweight children and adolescents in the United States has doubled and tripled, respectively (Baskin, Ard, Franklin, & Allison, 2005), while relative maximal oxygen consumption (VO_{2max}, a gold-standard measure of aerobic fitness) has decreased by 20% in adolescent females (Eisenmann & Malina, 2002). Many speculate that this trend is exacerbated by the decreased amount of time children spend being physically active at school, and correspondingly indicates that schools are the ideal environment for intervention (Donnelly et al., 2009; Hillman, Erickson, & Kramer, 2008). Such ideas have generated a strong interest in attempting to unravel the association between children's health and their endeavors at school, most notably academic achievement.

1.1. Children's aerobic fitness and cognition

A growing body of research has focused on the relationship between children's aerobic fitness and academic performance using field estimates and classroom evaluations in school. Although a causal relationship has not yet been established, a number of findings have demonstrated that more aerobically fit children perform better on standardized achievement tests and receive higher grades (see CDC, 2010, for review). Such findings have considerable implications for children's health and their accomplishments in school, especially given that early academic achievement often predicts future engagement and success. This relationship has been particularly well established for language and reading abilities. Superior skills (phonological awareness, word decoding, reading comprehension, and spelling) and early performance have been linked to better grades and greater reading engagement during high school (e.g., Cunningham & Stanovich, 1997; Lonigan, Burgess, & Anthony, 2000), as well as improved mastery of a second language (which many children also encounter in the school curriculum; Sparks, Patton, Ganschow, Humbach, & Javorsky, 2008).

* Corresponding author. Address: Department of Kinesiology & Community Health, University of Illinois, 317 Louise Freer Hall, 906 South Goodwin Avenue, Urbana, IL 61801, USA. Fax: +1 217 244 7322.

E-mail addresses: mscudde2@illinois.edu (M.R. Scudder), kfederme@illinois.edu (K.D. Federmeier), lraine2@illinois.edu (L.B. Raine), artur.direito@nihi.auckland.ac.nz (A. Direito), jboyd@crl.ucsd.edu (J.K. Boyd), chhillma@illinois.edu (C.H. Hillman).

[1] Present address: The University of Auckland, Auckland 1142, New Zealand.

[2] Present address: University of California San Diego, La Jolla, CA 92093, USA.

http://dx.doi.org/10.1016/j.bandc.2014.03.016
0278-2626/Published by Elsevier Inc.

However, most previous research investigating the relationship between fitness and academic achievement has relied exclusively on behavioral outcomes, which account for overt performance but, as end-state measures, summate across multiple cognitive processes that may change as a function of fitness. Thus, when higher fit children outperform their lower fit peers on tests like reading achievement, it is unclear whether this is due to factors such as motivation and attention-to-task, or because fitness has more direct effects on specific language subprocesses. For this reason, event-related brain potentials (ERPs) have played a critical role in research on fitness effects, as this method provides a time-sensitive record of the brain's response during cognitive engagement. Aspects of the ERP response "components" have been linked to specific cognitive processes, and those responses are multidimensional in nature, consisting of changes in amplitude, latency, and/or topography across groups and/or experimental conditions. Therefore, ERPs provide a set of functionally specific dependent measures that can help reveal the neural basis of fitness-related associations with academic achievement and cognitive processing in a way that is not possible with behavioral outcomes alone.

For example, recent investigations focusing on the P3b (Hillman, Buck, Themanson, Pontifex, & Castelli, 2009; Hillman, Castelli, & Buck, 2005; Pontifex et al., 2011) in children have successfully demonstrated an association between aerobic fitness and specific, core aspects of cognition. The P3b is elicited when people attend to or discriminate between stimuli (Polich & Kok, 1995). The amplitude and latency of this response are influenced by the availability of attentional resources and the time necessary to evaluate and update stimulus context (e.g., Donchin & Coles, 1998; Polich, 2007). P3b responses have been collected from higher and lower fit children using paradigms including oddball tasks (Hillman et al., 2005) and modified versions of the Eriksen flanker task (Eriksen & Eriksen, 1974; Hillman et al., 2009; Pontifex et al., 2011), each of which challenges participants to attend to specific stimuli, inhibit pre-potent tendencies or distracting information, and respond both quickly and accurately. In addition to greater response accuracy and shorter response times (RT), higher fit children exhibit larger P3b amplitude and shorter latency compared to those with lower fitness. These findings link better behavioral performance in higher fit children to differences in attention and inhibitory control, with higher fit children able to efficiently and effectively allocate greater attentional resources to support task performance (Hillman et al., 2005, 2009; Pontifex et al., 2011).

Accordingly, the functional capabilities of ERPs are well-suited for examining skills such as reading, which unfold rapidly and across time. In addition, as described next, ERPs provide functionally well-specified measures of core language processes; thus, with ERPs it is possible to look for relationships between fitness and specific language processing abilities. Therefore, the current study aims to extend research examining fitness and children's cognition by investigating ERP indices of meaning processing (the N400 component) and the analysis of language structure or grammar (frontal negativities and the P600).

1.2. The N400 component and N400 effects

The N400 is an ERP component that is elicited by words in all modalities, as well as other types of meaningful stimuli, and has been linked to the access of meaning information from long-term memory (e.g., Lau, Phillips, & Poeppel, 2008; see review in Kutas & Federmeier, 2011). In the thirty years since its discovery (Kutas & Hillyard, 1980), this response has been well-established as a reliable measure of semantic processing. As its name implies, the N400 is a negative-going ERP component observed approximately between 300 and 500 ms (peaking around 400 ms) after stimulus presentation, and is widely-distributed over the scalp with a topographic maximum over central and parietal regions (Holcomb, Coffey, & Neville, 1992). The N400 can be observed in children as young as 19 months (Friedrich & Friederici, 2004), and its latency and amplitude index different aspects of language abilities along with their changes across the lifespan. Across multiple studies and participant populations, evidence suggests that N400 latency decreases over the course of childhood (Hahne, Eckstein, & Friederici, 2004; Holcomb et al., 1992; Juottonen, Revonsuo, & Lang, 1996), as language abilities increase. The relationship between N400 latency and language fluency has also been documented in studies of bilingual participants, which allow a within participant comparison between languages with which these participants are more or less fluent. These studies show that the N400 peaks later in the less dominant language and that, more generally, shorter N400 latencies are associated with greater language proficiency (e.g., Moreno & Kutas, 2005).

The amplitude of the N400 varies with the amount of lexico-semantic activation elicited by an incoming word (see Kutas & Federmeier, 2011, for review). N400 amplitude is increased for words with more orthographic neighbors (the number of words known to the reader that share all but one letter in common with the stimulus), higher neighborhood frequency (the average frequency of the orthographic neighbors), and more lexical associates (Holcomb, Grainger, & O'Rourke, 2002; Laszlo & Federmeier, 2007, 2011; Laszlo & Plaut, 2012). N400 amplitudes are also larger for concrete than abstract words (see review in Lee & Federmeier, 2012) and for known than unknown words, both in toddlers (Mills, Coffey-Corina, & Neville, 1993, 1997; Mills, Conboy, & Paton, 2005) and adults learning a second language (McLaughlin, Osterhout, & Kim, 2004). Thus, all else being equal, words that make contact with a larger set of lexical items and/or semantic features elicit larger N400 responses. Correspondingly, N400 component amplitude provides a measure of language ability. For example, in school-age children, Coch and Holcomb (2003) found larger N400 amplitudes in high compared to low ability readers across a range of different word types presented in isolation.

Relative to its "baseline" level, the amount of N400 activity a given word elicits can be reduced by factors that ease lexico-semantic processing. N400 amplitudes are reduced by repetition (and similarly, by higher word frequency when words are presented in isolation) and when context information (of many types) renders words more plausible. For example, N400 amplitudes are smaller for "*shoes*" than for "*songs*" in the sentence context "*You wear shoes/songs on your feet*" (see Kutas & Federmeier, 2011 for review). This difference between a word in an incongruent versus congruent context is typically referred to as the "N400 effect". The size of the N400 effect, independent of overall N400 amplitude, provides a measure of how well the available context information is being used during comprehension. As such, N400 effect size has been used to document changes in the ability to make use of sentence context information with normal aging (Wlotko, Lee, & Federmeier, 2010) and schizophrenia (Sitnikova, Salisbury, Kuperberg, & Holcomb, 2002). In the literature on language learning, pictures have been used as context, with the N400 effect to matching versus mismatching words then constituting an implicit vocabulary measure (e.g., Byrne, Dywan, & Connolly, 1995; Byrne et al., 1999; but see Henderson, Baseler, Clarke, Watson, & Snowling, 2011, for a failure to replicate this correlation with vocabulary).

In summary, measurements of the N400 can provide multiple, complementary indices of language comprehension abilities. The size and timing of the N400 component reflect, respectively, the richness of the mental lexicon and the speed with which it can be accessed. Assessments of the N400 effect (facilitation for plausible relative to unexpected/incongruent words) in a particular experimental context can additionally speak to a person's ability to use that type of context information to facilitate word processing online.

1.3. Anterior negativities and the P600 component

Other aspects of language processing are indexed by different parts of the ERP waveform. Violations of language structure have been associated with both frontally-distributed negativities between 100 and 500 ms (ELAN and LAN: (Early) Left Anterior Negativity), and a later (after 500 ms) posterior positivity (P600). Violations of word category (e.g., *"The man admired a sketch of the landscape."* versus *"The man admired Don's of sketch the landscape."*; Neville, Nicol, Barss, Forster, & Garrett, 1991) have been associated with ELANs in several studies, although the functional nature of this response remains controversial (for a review see Friederici & Kotz, 2003, and see Dikker, Rabagliati, & Pylkkänen, 2009, for an alternative perspective). Importantly for the present study, which uses 9–10 year old children, developmental investigations of the ELAN have suggested that it matures later than other language-related ERP responses and may not reliably occur in response to syntactic violations until the teenage years (e.g., Hahne et al., 2004).

In contrast, it has been found that children as young as 36 months elicit a broadly distributed positive component to auditory verb tense violations (Silva-Pereyra, Rivera-Gaxiola, & Kuhl, 2005), with adult-like P600 responses seen in children by ages 7–8 (Atchley et al., 2006; Hahne et al., 2004). The precise functional significance of the P600 continues to be debated, but it is reliably elicited by sentences containing syntactic violations as well as sentences with less-preferred or difficult syntactic structures, and has been associated with attentionally-demanding and effortful processes of integration, reanalysis, and/or repair (Brouwer, Fitz, & Hoeks, 2012; Friederici, 2002; Kuperberg, 2007). Given the literature showing fitness effects on the P3b, it is notable that some work has pointed to important similarities between the P600 and the P3b (but see Frisch, Kotz, Von Cramon, & Friederici, 2003, for data also highlighting differences between these responses). Coulson, King, and Kutas (1998) asked participants to read grammatical sentences mixed with those containing morphosyntactic violations, with the violations either improbable (20% violations) or probable (80% violations) within a given block of trials. Consistent with typical P3b patterns, larger P600s were observed for more salient stimuli (pronoun versus verb agreement violations in the current example), and for trials with lower probability. The authors suggested that, similar to interpretations of the P3b, the P600 may reflect comprehenders' ability to recognize ungrammatical events and adjust their expectations (i.e., update context) accordingly.

1.4. The current study

Investigations focusing on the relationship between aerobic fitness and the N400/P600 in humans are limited, with no published studies examining this relationship among children. Results from one study in adults found that aerobic fitness had no influence on the N400 (Magnié et al., 2000). However, given that reading and other aspects of language processing are undergoing critical development during childhood, this may be an important period of the lifespan in which fitness can be influential and thus when measurable differences in the N400 and P600 linked to fitness may be more likely to occur. Accordingly, the study described herein focuses on determining whether the cognitive benefits associated with aerobic fitness in children extend to neuroelectric measures of language processing. ERPs were recorded in higher and lower fit children as they read visually presented sentences with a 50% chance of containing a semantic anomaly or an unusual word order. To encourage attention to the sentences, children were asked to indicate whether the sentences contained a "mistake" by pressing a corresponding button on a response pad. Children were also tested on offline measures of academic achievement using the Wide Range Achievement Test 3rd edition (WRAT3; Wilkinson, 1993).

ERP responses were examined to the same, critical, mid-sentence target words in the three contexts or trial types: congruent, semantically anomalous, and out of order. Adult-like responses to these words would include an N400 component for each trial type, but the N400 should be reduced in size for words in congruent contexts compared to those resulting in a semantic anomaly (i.e., there should be an N400 effect). Larger N400 amplitude is also predicted for words occurring out of order – wherein a noun appears in a sentence position that would normally have a verb – as they contain less contextual support (Neville et al., 1991). Words out of order would also be expected to elicit a P600 if children's reading processes are adult-like. If the superior reading achievement previously reported in higher fit children reflects an effect of fitness on critical subcomponent processes of online language comprehension, one may expect that higher fit children would demonstrate greater N400 amplitude and/or decreased N400 component latency to reveal greater word knowledge and speed of lexical access, respectively. Better ability to make use of semantic context information would result in an enhanced N400 effect (more facilitation for congruent words compared to violations), and better ability to detect and evaluate grammatical information would result in enhanced P600 effects (more positive responses to word order violations than congruent words). Such findings would further support the growing body of work demonstrating that greater aerobic fitness is associated with improved cognitive performance (Hillman et al., 2008), and extend such findings to neural processes critical for language processing and reading achievement.

2. Methods

2.1. Participants

Forty-six preadolescent children (23 higher fit: 11 M/12F; 23 lower fit: 11 M/12F), ages 9–10 years, were recruited from local communities neighboring the University of Illinois at Urbana-Champaign campus. Demographic information for participants may be found in Table 1. Written assent and informed consent were collected from participants and their legal guardians in accordance with the Institutional Review Board. Prior to testing, guardians completed a health history and demographics questionnaire and the Pre-Participation Health Screening (HALO Research Group, 2010), which is designed to ensure participants are free of pre-existing conditions that could be exacerbated by physical exercise.

Table 1
Mean (SD) values for participant demographics.

Measure	Lower fit participants	Higher fit participants
n	23 (12 female)	23 (12 female)
Age (years)	10.0 (0.6)	9.9 (0.7)
ADHD	36.9 (22.5)	38.2 (28.9)
Socioeconomic Status (SES)	2.2 (0.7)	2.3 (0.8)
K-BIT Composite (IQ)	115.0 (10.2)	120.4 (10.6)
BMI (kg/m^2)*	21.0 (3.8)	15.8 (1.3)
BMI percentile*	33.0 (4.3)	80.0 (4.7)
VO_{2max} (ml/kg/min)*	37.2 (4.7)	51.5 (5.0)
VO_{2max} percentile*	11.0 (7.2)	81.8 (6.8)
HR_{max} (bpm)	194.9 (11.4)	195.9 (8.6)
PA week/weekend (hours)	2.4 (1.6)/2.6 (1.2)	2.4 (1.0)/2.8 (1.2)

Note: HR_{max} is the maximum heart rate achieved during the VO_{2max} (maximal oxygen consumption) test; PA week(end) is the average number of hours participants' report being physically active during the day.
* $p < .05$.

Guardians indicated that their children were free of attentional disorders (as indexed by the ADHD Rating Scale IV; DuPaul, Power, Anastopoulos, & Reid, 1998) and neurological diseases, had normal or corrected-to-normal vision, and were native English speakers raised in an English speaking household. Socioeconomic status was calculated using a trichotomous index based on: (1) highest level of education obtained by the mother and father, (2) number of parents who worked full time, and (3) participation in a free or reduced-price lunch program at school (Birnbaum et al., 2002). Pubertal timing data were used to ensure that all participants were prepubescent at the time of testing ($\leqslant$2 on a 5 point scale) using the modified Tanner Staging System (Tanner, 1962; Taylor et al., 2001). Children scoring at or above the 70th or below the 30th percentile for aerobic fitness (Shvartz & Reibold, 1990) were included in the study and placed into higher and lower fitness groups, respectively.[3]

2.2. Aerobic fitness

VO_{2max} was measured using a computerized indirect calorimetry system (ParvoMedics True Max 2400, Sandy, UT) during a modified Balke protocol (ACSM, 2010) conducted on a motor-driven treadmill. Participants were instructed to walk for 2 min before jogging at a constant and comfortable speed, with 2.5% increases in grade every 2 min until volitional exhaustion. Averages for oxygen uptake and respiratory exchange ratio (RER) were assessed every 20 s. A Polar heart rate (HR) monitor (Model A1, Polar Electro, Finland) measured HR throughout the test, and ratings of perceived exertion (RPE) were assessed every 2 min using the children's OMNI Scale (Utter, Robertson, Nieman, & Kang, 2002). VO_{2max} was expressed in ml/kg/min (milliliters of oxygen consumed per kilogram of body weight per minute) and was evidenced by a minimum of 2 of the following 4 criteria: (a) a plateau in oxygen consumption corresponding to an increase of less than 2 ml/kg/min despite an increase in workload; (b) RER $\geqslant$ 1.0 (Bar-Or, 1983); (c) a peak HR $\geqslant$ 185 beats per min (bpm; ACSM, 2010) and a HR plateau (Freedson & Goodman, 1993); and/or (d) RPE $\geqslant$ 8 (Utter et al., 2002).

2.3. Offline academic achievement

Reading, spelling, and arithmetic achievement were determined using the WRAT3, which has been age standardized and strongly correlated with the California Achievement Test and Stanford Achievement Test (Wilkinson, 1993). The WRAT3 includes two separate versions (blue and tan), which contain separate words and problems to enable researchers to measure performance at two different time points. In the current study, only one version was needed, and therefore the trained laboratory staff administered the blue form to each participant. The reading portion instructs participants to correctly pronounce as many words as possible from a list that becomes progressively more difficult. Similarly, the spelling and arithmetic portions instruct individuals to spell and solve increasingly more difficult words and problems, respectively.

2.4. Online sentence processing task

The sentences used in the current paradigm (see Appendix A) were presented using Neuroscan STIM[2] software (Compumedics, Charlotte, NC). The paradigm was originally developed for use in the auditory modality with preliterate children, and has been previously used with adults[4] (Aydelott et al., 2006). A total of 80 sentences were available in three different forms: congruent, semantic violation, and word order (i.e., syntactic) violation. Two congruent sentence examples are provided below with the target lexical item underlined:

Sentence A – "*You wear <u>shoes</u> on your feet*."
Sentence B – "*At school we sing <u>songs</u> and dance*."

Semantic violations were constructed by shuffling target words across sentences, such that the same lexical items became semantically anomalous continuations of other sentence contexts. For instance, "*shoes*," the congruent target word for Sentence A above, creates a semantic violation when placed into the target position of Sentence B. Lexically-controlled violations of canonical phrase structure (which we will refer to as "syntactic violations") were then created by reversing the order of the verb and its object noun (congruent: SVO – subject/verb/object; word order violation: SOV – subject/object/verb), as demonstrated below:

Sentence examples
Congruent – "*You wear <u>shoes</u> on your feet*."
Semantic – "*At school we sing <u>shoes</u> and dance*."
Syntactic – "*You <u>shoes</u> wear on your feet*."

Thus, across the experiment, the lexical items used as target words remained constant. Characteristics of the target words (number of letters: *M* = 5.1, SD = 1.7; written frequency: *M* = 99.2, SD = 171.9, Kuĉera & Francis, 1967; age of acquisition: *M* = 222.0, SD = 40.5; familiarity: *M* = 578.0, SD = 40.0) were gathered from the MRC Psycholinguistic Database (see Coltheart, 1981; Wilson, 1988, for unit descriptions). Orthographic neighborhood size of the target words (*M* = 8.3, SD = 6.4 neighbors) was calculated using Wuggy (Keuleers & Brysbaert, 2010).

Two master sentence lists were created, randomized, and counterbalanced across participants using all eighty congruent sentences. The violation versions of each sentence were evenly divided and counterbalanced across the lists for a total of forty semantic and syntactic violations in each list. Congruent versions of all sentences (80 total) also appeared in each list.

The order of appearance for each sentence version (congruent/violation or violation/congruent) was counterbalanced across lists. That is, in master list 1 if the congruent version of a particular sentence appeared in the first half of the list, it would appear in the second half of master list 2. Violation versions, either semantic or syntactic, would then appear in opposite halves of each master list accordingly. The 160 total sentences were divided into 8 separate blocks comprised of 20 sentences each (10 congruent, 5 each semantic and syntactic). Sentences ranged from 5 to 13 words long (*M* = 8.08, SD = 1.57 words) and were presented one word at a time on an LCD computer monitor at a distance of one meter (0.65° visual angle).

All sentences began with a fixation cross that was subsequently followed by individual words presented in white on a black background. Children were asked to monitor for semantic and syntactic violations, which were referred to as "mistakes". Following the end of each sentence, which was denoted with a period accompanying the last word, a question (i.e., "Were there any mistakes?") was presented on the screen asking participants to evaluate what they had read. Participants were asked to respond as quickly and accu-

[3] A total of twenty participants were excluded from the study: seventeen qualified as moderately fit (i.e., 31st–69th percentile), one guardian scored their child above the cutoff for ADHD, one participant indicated that English was not the primary language spoken in the family household, and one was unable to finish testing due to scheduling conflicts.

[4] The Aydelott, Dick, and Mills (2006) study used only the congruent and semantic violation conditions. However, unpublished data also using the word order violations with college-aged adults shows the expected P600 effect beginning around 800 ms over posterior electrode sites.

rately as possible with a left button press (using a response pad) if the sentence did not contain a mistake, or a right button press if the sentence did contain a mistake. The fixation cross and word stimuli were presented for 750 ms with a 250 ms inter-stimulus-interval (ISI). An additional 750 ms was added to the ISI between the last word of the sentence and the question to help denote the end of the sentence. The question was displayed for 6 s or until the participant answered, at which point the question disappeared and the screen turned black. Participants were given a 6250 ms response window, and the next fixation cross appeared after 7 s.

2.5. ERP recording

During sentence reading, electroencephalographic (EEG) activity was recorded from 64 electrode sites (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) and referenced to averaged mastoids (M1, M2), with AFz serving as the ground electrode. Electrodes were arranged in an extended montage based on the International 10–10 system (Chatrian, Lettich, & Nelson, 1985) using a Neuroscan Quik-Cap (Compumedics, Charlotte, NC). Electrooculographic (EOG) activity was recorded using additional electrodes placed above and below the left eye and on the outer canthus of each eye. All impedances were <10 kΩ. Continuous data were filtered with a DC to 70 Hz bandpass filter and a 60 Hz notch filter, amplified 500× using a Neuroscan SynAmps amplifier (Compumedics, Charlotte, NC), and sampled at 500 Hz. Offline data reduction included EOG correction using a spatial filter (Compumedics Neuroscan, 2003), and any trials containing an artifact exceeding ±75 μV were automatically rejected. Stimulus-locked epochs were created from −100 to 1000 ms around the target-word stimuli, and averages were constructed with a minimum of 25 trials. All averages consisted of epochs from the target word stimuli for each sentence. Baseline correction was conducted using the 100 ms pre-stimulus period, and data were filtered using a 30 Hz low pass and 0.01 Hz high pass filter.

2.6. Procedure

Participants completed 2 h of testing on two separate days and received $10/h remuneration. These days were scheduled approximately 1 week apart and occurred at the same time of day. Day 1 began with participants and guardians completing informed assent and consent, respectively. Hand dominance, intelligence quotient, and academic achievement were determined for each participant using the Edinburgh Handedness Inventory (Oldfield, 1971), Kaufman Brief Intelligence Test 2 (K-BIT2; Kaufman & Kaufman, 2004), and WRAT3 respectively. Next, participants were given the reading-task instructions, made explicitly aware of the two different forms of mistakes, and completed 20 practice sentences with an identical format to those which would be encountered on the second visit. Practice sentences were generated from an additional sentence list and were not included in either master list. Lastly, height and weight were recorded using a stadiometer and a Tanita WB-300 Plus digital scale before completing the VO_{2max} test.

On Day 2, participants were fitted with an electrode cap for neuroelectric measurement and seated in a sound-attenuated room. Each participant was read task instructions again and received a shorter eight sentence practice block consisting of 4 congruent, 2 semantic, and 2 syntactic trials before testing began. Participants were afforded the opportunity to ask questions and were given 1 min of rest between each block of sentences. After the final block, the electrode cap was removed and participants were briefed on the purpose of the experiment.

2.7. Statistical analysis

All statistical procedures were computed and variables of interest were analyzed using SPSS v.19 (IBM Corp., Armonk, NY). Analyses with three or more within-subjects levels report p-values after Greenhouse–Geisser correction for violations of sphericity. Significance levels were set at $p = .05$, and post hoc comparisons were conducted using Bonferroni correction. Cohen's d is reported to indicate effect size, and standard errors are reported in addition to the group means. To determine if the sentence presentation order (congruent 1st – violation 2nd, or vice versa) of each master list had an influence on children's behavior/ERPs, the list that they completed was incorporated as a factor (i.e., "Order") in each of the corresponding analyses. The results revealed no significant effects involving Order in any analysis (all p's $\geqslant .11$), so these effects are not reported in detail. Effects of electrode site are also not reported, as they are not of theoretical significance to the aims of the study.

3. Results

3.1. Academic achievement

Academic achievement was evaluated using separate independent-samples t-tests to investigate fitness (higher fit versus lower fit) differences for reading, spelling, and arithmetic. As can be seen in Fig. 1A, analysis revealed that higher fit children ($M = 123.1 \pm 2.8$ standardized score) had greater reading achievement scores compared to lower fit children ($M = 112.9 \pm 2.1$ standardized score), $t(44) = 2.8$, $p = .007$, $d = 0.8$. A similar finding was observed for spelling (higher fit: $M = 117.6 \pm 2.6$ standardized score; lower fit: $M = 108.0 \pm 3.0$ standardized score), $t(44) = 2.4$, $p = .02$, $d = 0.7$. No fitness differences were observed for arithmetic achievement, $t(44) = 1.4$, $p = .18$, $d = 0.4$. Pearson product-moment correlations between WRAT3 performance and the other dependent cognitive measures are given in Table 2. Although spelling performance was associated with congruent trial accuracy, $r = .35$, and arithmetic with syntactic trial accuracy, $r = .34$, WRAT3 measures were not correlated with any neuroelectric variables, $r \leqslant .26$, $p \geqslant .07$.

3.2. Sentence task RT and accuracy

The behavioral task was designed to ensure that children were paying attention to the sentences and could comprehend them. RT and accuracy were examined using a repeated measures multivariate analysis of variance (MANOVA) with Order (master list 1 and 2) and Fitness (higher and lower) entered as between subjects variables, and Trial Type (congruent, semantic, and syntactic) as a within subjects variable.[5] For RT, there were effects of Fitness, $F(1,42) = 5.8$, $p = .02$, $\eta^2 = .12$, and Trial Type, $F(2,82.9) = 30.6$, $p < .001$, $\eta^2 = .42$. Higher fit children ($M = 1021.8 \pm 63.4$ ms) had overall shorter RTs compared to lower fit children ($M = 1250.9 \pm 68.5$ ms), $d = 0.7$ (see Fig. 1B). The effect of Trial Type arose because syntactic violation trials ($M = 976.0 \pm 45.3$ ms) elicited shorter RTs compared to semantic violation ($M = 1173.3 \pm 52.3$ ms),

[5] Response button assignment was not counterbalanced across participants, as ERP components were the primary outcome measure, yet independent samples t-tests indicated that performance did not differ between left-handed ($n = 7$; 4 higher fit, 3 lower fit) and right-handed individuals, suggesting button assignment had little influence on the behavioral outcomes. Similarly, distributional ERP differences in these individuals were not of interest or concern, as the analyses revealed that the N400 effect and maximal amplitude were centered over midline sites in children as a whole (see Fig. 3A and B). Further, among just left-handed participants, N400 amplitude/latency differences between higher (−6.8 ± 1.9 μV; 382.8 ± 11.8 ms) and lower fit (−2.9 ± 1.7 μV; 433.2 ± 14.1 ms) children were consistent with the overall patterns reported.

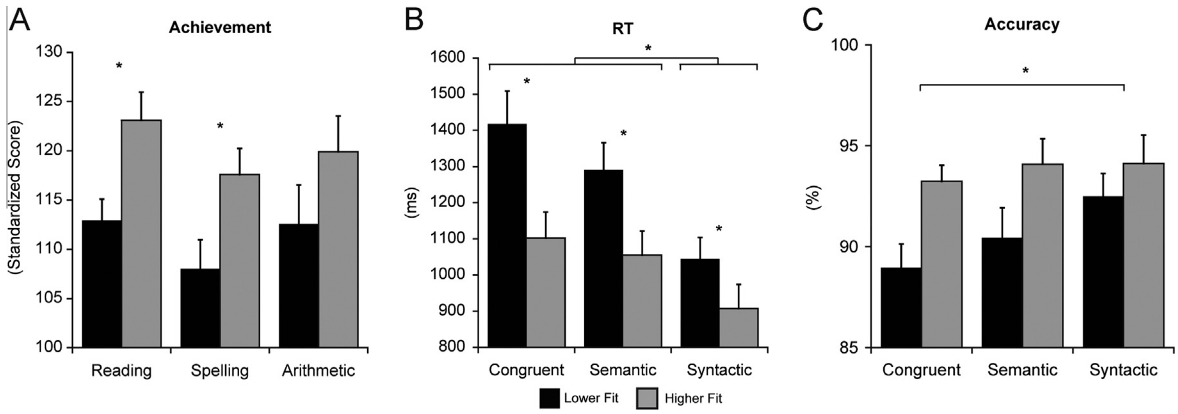

Fig. 1. (A) Higher fit children outperformed lower fit children on tests of reading and spelling achievement. (B) Higher fit children exhibited overall shorter response time (RT). Further, shorter RT was observed for syntactic compared to congruent and semantic sentences across all children. (C) Overall performance was high in both groups of children. Analysis also revealed that higher fit children performed more accurately across all sentence types. Error bars represent standard error. *$p < .05$ between fitness groups.

Table 2
Pearson correlations between WRAT3 academic achievement and other cognitive measures.

Measure	Reading	Spelling	Arithmetic
Congruent (RT/accuracy)	(−.05/.15)	(−.17/.35*)	(−.21/.26)
Semantic (RT/accuracy)	(−.11/.15)	(−.15/.13)	(−.22/.15)
Syntactic (RT/accuracy)	(−.04/.19)	(−.05/.29)	(−.12/.34*)
Congruent N400 (latency/amplitude)	(−.10/.05)	(−.23/.08)	(−.20/.22)
Semantic N400 (latency/amplitude)	(−.20/−.05)	(−.27/−.08)	(−.16/.13)
Syntactic N400 (latency/amplitude)	(−.14/−.09)	(.04/−.05)	(−.07/.03)
N400 effect (semantic/syntactic)	(−.13/−.17)	(−.21/−.15)	(−.05/−.20)
P600 effect	(−.05)	(−.03)	(−.19)

Note: Children who performed better on academic tests of spelling and arithmetic demonstrated greater accuracy for congruent and syntactic trials, respectively.
* $p < .05$.

$t(45) = 5.6$, $p < .001$, $d = 0.6$, and congruent ($M = 1259.8 \pm 62.2$ ms) trials, $t(45) = 7.1$, $p < .001$, $d = 0.8$. Semantic violation and congruent trial RTs did not differ, $t(45) = 2.4$, $p = .023$, $d = 0.2$. As for accuracy, there was a main effect of Fitness $F(1,42) = 5.4$, $p = .025$, $\eta^2 = .11$, indicating higher fit children ($M = 93.8 \pm 1.0\%$) had overall greater reading accuracy than lower fit children ($M = 90.6 \pm 0.9\%$; see Fig. 1C). Importantly, overall accuracy for both groups was very high, showing that all children could comprehend the sentences and detect both semantic and syntactic errors.

3.3. ERP measures

3.3.1. N400 latency

N400 local peak latency (Luck, 2005) was assessed during the 300–500 ms time interval and characterized using 25 electrode sites (Fz, FCz, Cz, CPz, Pz, F3/1/2/4, FC3/1/2/4, C3/1/2/4, CP3/1/2/4, P3/1/2/4) entered into an Order × Fitness × Trial Type × Electrode repeated measures MANOVA. There was a main effect of Fitness, $F(1,42) = 14.7$, $p < .001$, $\eta^2 = .26$, that revealed higher fit children ($M = 392.7 \pm 5.4$ ms) had shorter N400 latencies compared to their lower fit peers ($M = 422.0 \pm 5.4$ ms), irrespective of sentential context, $d = 1.2$ (i.e., whether the target lexical item resulted in a violation or not; see Fig. 2).

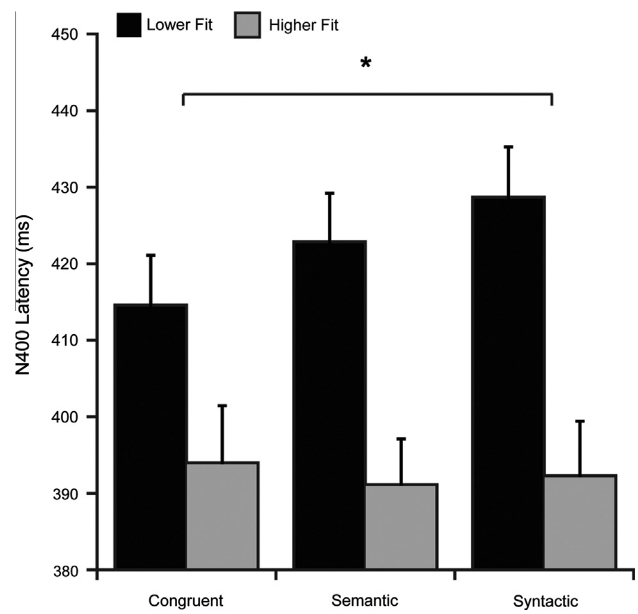

Fig. 2. Findings revealed that higher fit children had shorter N400 latency across all trial types compared to lower fit children. Error bars represent standard error. *$p < .05$ between fitness groups.

3.3.2. N400 effect and component amplitude

The results of the N400 latency analysis were used to select a time window for assessment of the N400 component and effect amplitude that was centered around both fitness groups' peak latencies, while still being narrow enough to reduce the possibility of overlap between the N400 and other components, such as the P600 (Coulson et al., 1998). As the peaks were 30 ms apart, we chose a time window that began 30 ms before the earlier peak and extended to 30 ms after the later peak – hence, 360–450 ms. To choose appropriate channels for the analysis of N400 component amplitude, we first assessed the size and topography of the N400 effect taken from the average difference waveforms (difference between violation and congruent trials), using the same 25 electrode sites described above. These measures were subjected to an Order × Fitness × Effect Type (semantic violation effect, syntactic violation effect) × Laterality (left lateral, left medial, midline, right medial, right lateral) × Anteriority (frontal, frontal central, central, central parietal, parietal) repeated measures MANOVA. A main effect of Anteriority, $F(1.5,61.8) = 6.0$, $p = .01$, $\eta^2 = .12$, indicated that children exhibited a typical N400 effect distribution, with maximal differences seen over central to parietal regions (see Fig. 3A). There was no main effect of Fitness, $F(1,42) = 1.6$, $p = .21$, $\eta^2 = .04$, and Fitness did not interact with Effect Type, $F(1,42) = 0.3$, $p = .59$, $\eta^2 < .01$, or any distributional variable, $F \leqslant 2.1$, $p \geqslant .12$, $\eta^2 \leqslant .05$.

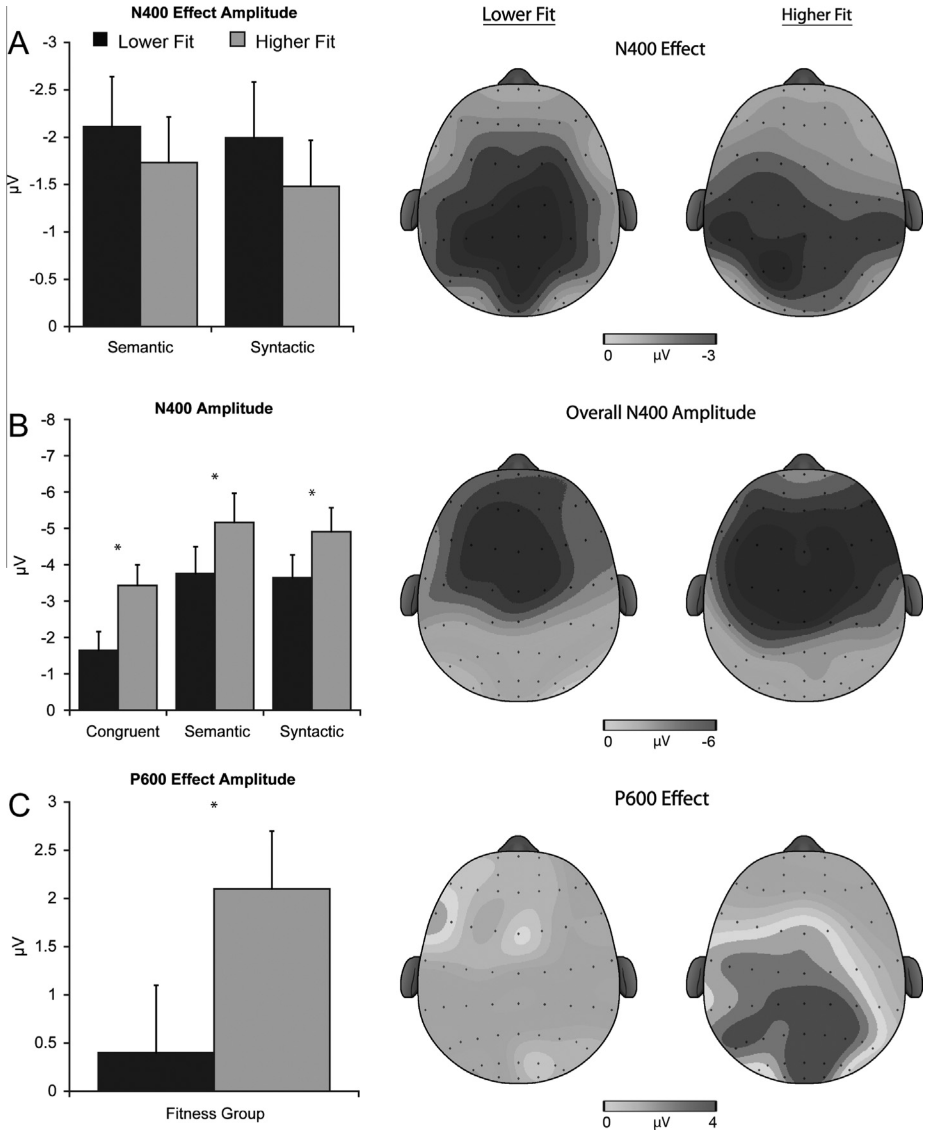

Fig. 3. (A) Comparable N400 effects were observed in higher and lower fit children over medial electrodes from central to parietal regions. (B) N400 amplitude was larger in response to violations in all children. Additionally, fitness differences were apparent across all trial types with higher fit children eliciting greater N400 amplitude. Topographic maps in (A) and (B) are collapsed across condition and represent the 360–450 ms time range used for analysis. (C) Syntactic violations elicited a P600 effect that was prominent in higher fit children. Error bars represent standard error. $^{*}p < .05$ between fitness groups.

N400 component amplitude was therefore assessed over the nine medial central parietal electrode sites where the N400 effect was maximal in the previous topographic analysis. Data were subjected to a repeated measures MANOVA with Order, Fitness, Trial Type, and Electrode (Cz, CPz, Pz, C1/2, CP1/2, and P1/2) as factors. There were main effects of both Fitness, $F(1,42) = 7.0$, $p = .012$, $\eta^2 = .14$, and Trial Type, $F(2.0,83.8) = 20.6$, $p < .001$, $\eta^2 = .33$, but no interaction between the two.[6] Higher fit children ($M = -6.3 \pm 0.7\ \mu V$) exhibited greater N400 amplitude across all trial types compared to those who were lower fit ($M = -3.9 \pm 0.7\ \mu V$), $d = 0.8$ (see Figs. 3B, 4, and 5A). Additionally, semantic ($M = -6.0 \pm 0.6\ \mu V$) and syntactic violations ($M = -5.8 \pm 0.5\ \mu V$) both resulted in greater N400 amplitude compared to congruent trials ($M = -3.3 \pm 0.5\ \mu V$), $t(45) \geqslant 5.4$, $p \leqslant .001$, $d \geqslant 0.7$, but did not differ from one another, $t(45) = 0.1$, $p = .89$, $d = 0.1$.

3.3.3. P600 effect amplitude

Whereas the N400 is part of the normal electrophysiological response to words, and therefore should be present in all conditions, the P600 is predicted to occur only in response to syntactic violations.[7] Therefore, we examined the size and topography of the P600 effect (difference between syntactic violation and congruent trials), following the same analytical procedures used to examine the N400 effect.[8] Mean amplitudes taken from 600 to 900 ms were

[6] The N400 effect was analyzed first because its topography in young adults is well-characterized, and we wanted to ascertain whether our sample also showed the expected central/parietal topographic maximum (e.g., Holcomb et al., 1992). Given that we obtained the typical topography, we then conducted more focused tests over the central–parietal sites where the N400 exhibited its topographic maximum amplitude. However, the same effect on N400 amplitude is obtained even if analyses are conducted over all electrode sites: main effects of Fitness, $F(1,42) = 4.0$, $p = .04$, $\eta^2 = .10$, and Trial Type, $F(2.0,83.5) = 20.8$, $p < .001$, $\eta^2 = .33$.

[7] Although semantic violations are sometimes associated with positivities that follow the N400, these are not generally classified as P600 responses and their relationship to the P600 seen to syntactic violations is unclear (Van Petten & Luka, 2012).

[8] It was not possible to accurately measure peak latency for the P600 because the component manifested as an extended effect.

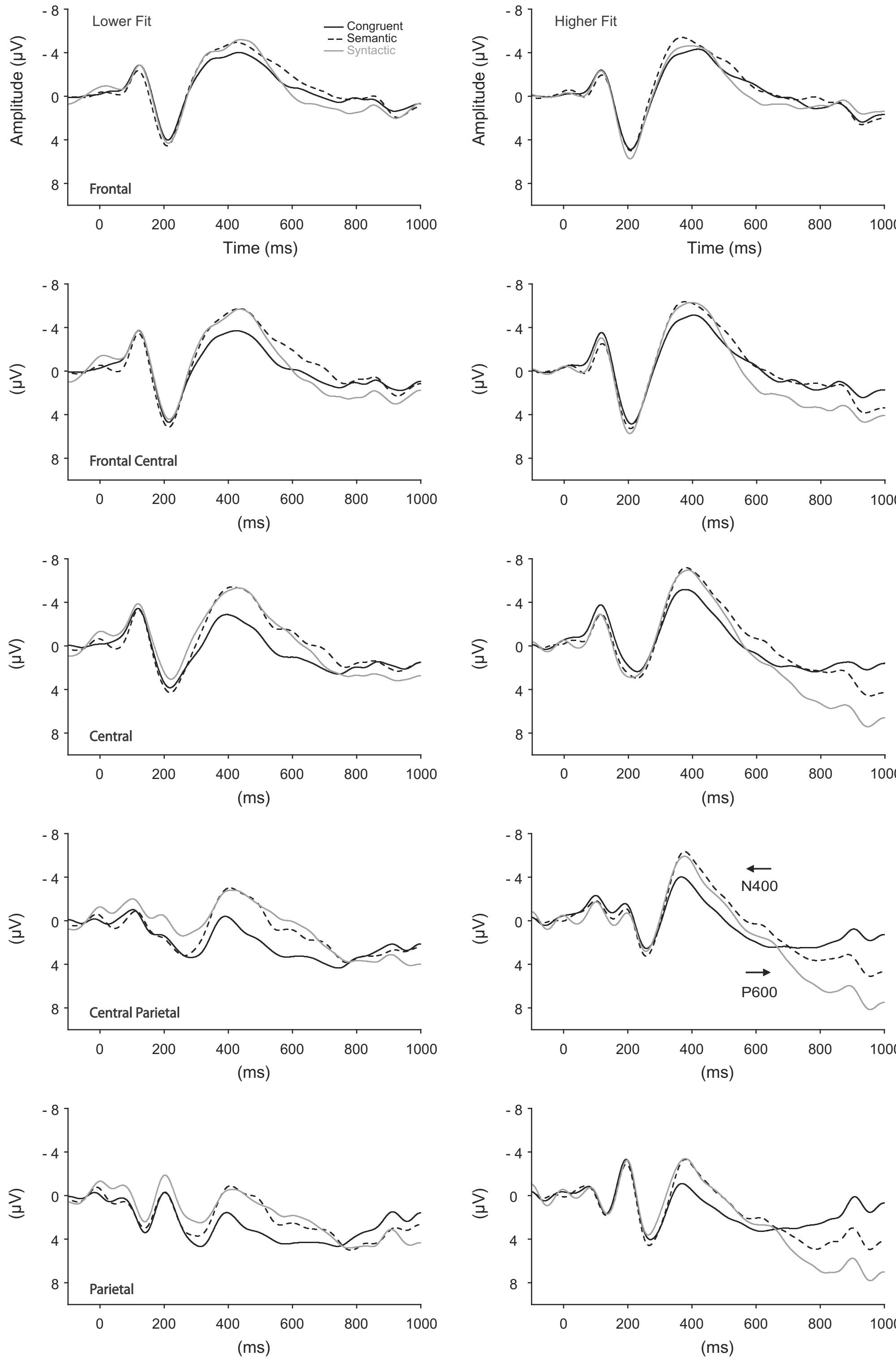

Fig. 4. Grand average waveforms for lower fit (left) and higher fit (right) children are depicted from frontal to parietal regions, collapsed across electrode within each region.

subjected to an Order × Fitness × Laterality × Anteriority repeated measures MANOVA. Analysis revealed a Fitness × Anteriority, $F(1.5,62.1) = 3.9$, $p = .037$, $\eta^2 = .09$, interaction indicating that higher fit participants ($M = 3.0 \pm 0.8$ μV) exhibited a larger P600 effect over the central parietal region compared to lower fit children ($M = -0.4 \pm 1.0$ μV), $t(44) = 2.6$, $p = .01$, $d = 0.8$ (see Figs. 4 and 5B), and that this effect approached significance over central and parietal regions, $t(44) \geqslant 1.9$, $p \leqslant .064$, $d \geqslant 0.6$ (see Fig. 3C). No fitness differences were observed over frontal or frontal central regions, $t(44) \leqslant 0.8$, $p \geqslant .41$, $d \leqslant 0.3$. To follow up on the Fitness × Anteriority interaction, mean amplitude over central parietal sites (the same 9 electrodes used for the N400 analysis: Cz, CPz, Pz, C1/2, CP1/2, and P1/2) was entered into a paired-samples t-test for each fitness group. These tests confirmed that syntactic trials engendered a P600 compared to congruent trials in higher fit individuals ($p = .001$), yet this effect was not observed in the lower fit group ($p = .88$). For

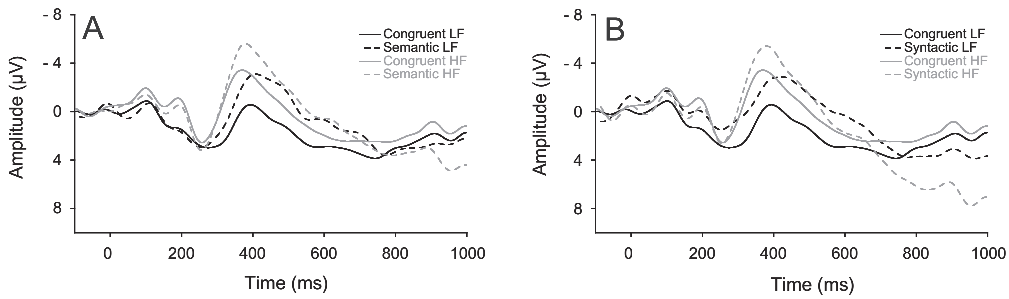

Fig. 5. Grand average waveforms are displayed for both lower and higher fit children from the 9-electrode (Cz, CPz, Pz, C1/2, CP1/2, and P1/2) region of interest. A comparison of semantic violations are shown in (A), and syntactic violations in (B).

completeness, a pairwise comparison of congruent and semantic violation trials was also included; as expected, these trials did not differ in either group (p's $\geqslant$ .32).

4. Discussion

The current findings replicate and extend prior results demonstrating a positive association between academic achievement and children's aerobic fitness (Castelli, Hillman, Buck, & Erwin, 2007; Dwyer, Sallis, Blizzard, Lazarus, & Dean, 2001; Roberts, Freed, & McCarthy, 2010). More aerobically fit children, as determined by relative VO_{2max}, scored higher on academic tests of reading and spelling, which measured their ability to correctly pronounce and spell progressively more difficult words. Of critical interest for this study then, was to use electrophysiological measures to uncover the potential mechanisms underlying these correlations between fitness and offline measures of language ability. To accomplish this aim, we measured neuroelectric activity as children read sentences, some of which contained a semantically anomalous word or an unusual word order. In particular, we examined whether higher levels of fitness were associated with the modulation of ERP responses linked to specific aspects of online language processing including meaning processing (the N400 component) and the analysis of language structure or grammar (the P600). ERP measures replicated well-established findings in the language processing literature. All target words elicited clear N400 responses over central and parietal regions of the scalp. Lexically unexpected words (both semantic and syntactic violations) elicited the large N400s typical of words without contextual support (Kutas & Hillyard, 1980; Neville et al., 1991), whereas congruent targets elicited facilitated (smaller) N400s (see Kutas & Federmeier, 2011 for review). Syntactic violations also produced the expected later posterior positivity associated with language structure (P600). Critically, both of these responses were modulated by aerobic fitness.

4.1. Aerobic fitness and the N400

Fitness differences were observed for both the timing and the amplitude of the N400 component, which has been linked to the access of semantic information from long-term memory. Across all trial types, higher fit children exhibited shorter N400 latency compared to lower fit children. Shorter N400 latencies point to more mature neurocognitive development, as N400 responses have consistently been found to decrease in latency across childhood and into the teenage years (Hahne et al., 2004; Holcomb et al., 1992; Juottonen et al., 1996). Within individuals of a given age, N400 latencies have been shown to be reduced as a function of language expertise. For example, investigations of bilingual speakers have shown that greater language proficiency (as determined by larger vocabulary), both across and within specific languages, is associated with shorter N400 latency (Moreno & Kutas, 2005). Thus, the N400 latency data show that higher fitness in children is associated with faster access to information about words and their meanings.

Higher fit children also elicited larger amplitude N400 responses than did lower fit children. Because words that share features or are associated with one another tend to become active in parallel (Laszlo & Plaut, 2012), overall N400 amplitude is an important indicator of the richness of the lexico-semantic network associated with a given word in a particular individual. Words with more lexical associates and greater neighborhood densities elicit larger N400 responses, both in and out of sentence contexts (e.g., Laszlo & Federmeier, 2008, 2011). In language learners, words that are known (versus unknown), elicit larger N400s (McLaughlin et al., 2004; Mills et al., 1993, 1997, 2005). Coch and Holcomb (2003) also found that first grade high ability readers elicited larger N400 amplitudes to single word stimuli than did low ability readers, although another study (Henderson et al., 2011) found that larger N400 amplitudes were associated with poorer ability to decode non-words in a sample of 8–10 year old children. Here, the overall larger amplitude N400 responses in higher fit children show that fitness is associated with enriched activation of long-term memory in response to words, especially taken in conjunction with the earlier N400 peak latencies seen in this group.

As already discussed, higher fit children achieved higher scores on offline measures of academic achievement related to language skills. The neuroelectric measures, however, were not directly correlated with the WRAT3 scores. As expected, these measures tap into at least partially independent skills that contribute to reading success. The tests in the WRAT3 assess explicit, word-form-related knowledge, such as the ability to spell words and to map orthographic forms onto phonological representations to allow correct pronunciation. The N400, in contrast, is an online measure that indexes early, relatively implicit, aspects of extracting meaning from words (see Kutas & Federmeier, 2011 for review). The N400 data thus provide a unique window into critical aspects of reading – namely, the ability to comprehend the meaning of printed words – that can be difficult to assess with standardized behavioral metrics alone, and that also seem to be augmented by higher levels of aerobic fitness in children.

Whereas overall N400 latency and amplitude differed across fitness groups, N400 effects, reflecting the amount of facilitation when a word is expected in its context relative to when it is unexpected, did not differ as a function of fitness in this study. Both groups demonstrated the expected N400 congruency effect, with reduced N400 amplitude to target words in the congruent condition compared to the same lexical items when they were semantically incongruent or presented in an unexpected order. Thus, both groups were clearly able to use context information to facilitate meaning processing for the congruent words. Some work has pointed to differences in N400 effect amplitude as a function of age and/or language proficiency (e.g., Friedrich & Friederici, 2006), but such differences appear to depend on the nature and difficulty of the manipulation (e.g., Byrne et al., 1995). To ensure that the sentences would be accessible to all readers for this first investigation into fitness effects on the neurophysiology of language, we used materials designed for very young children, which contained simple, well-known words and a very basic sentence structure. Accordingly, all children achieved high levels of accuracy (>90%) on the delayed behavioral classification task, showing that they could comprehend the sentences and detect semantic and syntactic violations, although higher fit children had shorter RTs and greater accuracy overall. Under these conditions, we also did not see effects of fitness on the ability to use context information to facilitate online word processing. However, it is of course possible that fitness-related differences on N400 congruency effects might emerge under conditions in which sentence processing was more taxing.

4.2. Aerobic fitness and the P600

Fitness effects were observed not only for semantic processing, in the form of earlier and larger N400 responses, but also for syntactic processing, on the P600. Higher fit children exhibited significantly larger P600 effects to words that were presented in a non-canonical order. The P600 is thought to arise from effortful processes involved in the integration or reanalysis of words, and has been observed in response to a variety of types of syntactic violations and sentences that engender syntactic processing difficulties (Brouwer et al., 2012; Friederici, 2002). Both fitness groups behaviorally identified these sentences as containing mistakes and both groups elicited larger N400s to these unexpected words. However, higher fit children appeared more capable of bringing online processes that have been associated with the analysis and/or repair of these errors, as reflected by the larger P600 effect. Given that prior research has highlighted important similarities and differences between the P3b and the P600 (e.g., Coulson et al., 1998), it is compelling that the effects of fitness seen in the present study pattern with the increased P3b amplitudes associated with fitness in prior work (Hillman et al., 2005, 2009; Pontifex et al., 2011). Despite evidence showing that elicitation of the P3b and P600 rely on at least partially dissociable neural networks (Frisch et al., 2003), the pattern of fitness-related ERP differences in children may suggest a link between these components and the neural/cognitive processes they embody. Thus, higher aerobic fitness levels in children may affect brain areas and mechanisms that mediate how (possibly domain-general) attention-demanding, categorization-like processes unfold in multiple types of tasks, including language comprehension.

4.3. Neural mechanisms underlying the impact of aerobic fitness

The impact of fitness on brain integrity is thought to partly arise from increased levels of neurotrophic factors, such as brain-derived neurotrophic factor (BDNF) and insulin-like growth factor (IGF), that support cell proliferation and differentiation (Vaynman & Gomez-Pinilla, 2005). In turn, the amplitude of anterior medial temporal lobe activity, an important source for the N400 component (McCarthy, Nobre, Bentin, & Spencer, 1995; Nobre & McCarthy, 1995), is correlated with neuronal density of the CA-1 region of the hippocampus (Grunwald, Lehnertz, Heinze, Helmstaedter, & Elger, 1998) and is reduced by antagonists for *N*-methyl-D-aspartate (NMDA) receptors, which are abundant in the hippocampus (Grunwald et al., 1999). Thus, it is conceivable that fitness effects on the development/integrity of hippocampal structures (Chaddock, Erickson, Prakash, Kim, et al., 2010), known to play a critical role in creating integrated, multimodal representations (Eichenbaum, Otto, & Cohen, 1994), would influence the strength and rate at which the brain is able to retrieve widely distributed representations across a multimodal long term memory system (Kutas & Federmeier, 2000). Similarly, accounts of greater P3b amplitude and the flexible regulation of attention in higher fit children appear to be related to larger volume of the basal ganglia (dorsal striatum; Chaddock, Erickson, Prakash, VanPatter, et al., 2010). Given the similarity between the P600 and the P3b components, as well as evidence pointing to a critical role for the basal ganglia in P600 production in particular (Frisch et al., 2003), it is possible that the underlying integrity of this brain region may account, in part, for the fitness effects observed herein.

4.4. Limitations and future directions

As with any cross-sectional study, the inherent nature of its design poses limitations. It is particularly important to acknowledge this with respect to the present study due to the fact that both groups differed not only on fitness level, but on academic achievement as well. Although this was expected and replicates previous results, it does limit the interpretation of the current ERP findings. As reported earlier, Coch and Holcomb (2003) have demonstrated that children with higher reading ability exhibit larger N400 amplitude compared to lower ability readers, much like the pattern observed involving higher and lower fit children. Despite the absence of any significant correlation between the ERP measures and the WRAT3, there is no way to determine if the significant relationship between fitness and N400 amplitude (or the other neuroelectric measures) is in fact causal. However, as it stands, these data are among the first to suggest an association between aerobic fitness and neuroelectric indices of children's language processing. Magnié et al. (2000) reported no such fitness relationship between 10 young adult elite cyclists and sedentary control subjects that read sentences with semantically congruous and incongruous endings, but they did report that overall N400 amplitude increased following a bout of maximal intensity aerobic exercise (but only for incongruous trials). Unfortunately, rest and exercise sessions (which were only separated by a maximal bout of exercise [i.e., a VO_{2max} test]) were not counterbalanced, and even though two separate sentence lists were utilized, this opens the possibility that the specific lexical items used to evoke the N400 may have influenced the results (in addition to the session order). Thus, a major strength of the current study includes the careful control of the lexical items used to elicit the N400 across the two master lists, as well as across participants. The relatively large number of children involved in the study, which were balanced across a number of demographic variables, including sex, was another highlight of the current investigation.

Accordingly, there is a strong need for future research encompassing fitness, exercise, and ERP indices of language processing among individuals of all ages. Converging evidence from such studies would help researchers further elucidate how maintaining an active lifestyle might help people achieve optimal cognitive functioning starting from an early age. This is precisely why current state and federal mandates (e.g., the No Child Left Behind Act) that allocate time *away* from physical activity in an attempt to improve children's scholastic performance may be negatively affecting the very processes they are designed to improve (Institute of Medicine of the National Academies, 2013). However, without being able to definitively explain how or why particular aspects of health (i.e., aerobic fitness) influence specific cognitive processes underlying school performance, changing current approaches will be hard fought. Given that language and reading skills are among the core predictors of future success and engagement in school, these areas deserve ample attention with respect to many different areas of children's physical well-being. Such information would be vital for school administrators and parents who are undoubtedly invested in improving the overall cognitive health of their children. For this reason, ERP measures, including the N400 and P600, are particularly well-suited to the goal of characterizing how fitness shapes cognition, as few other methods are capable of delineating multifaceted effects on specific cognitive subprocesses as they unfold over time.

5. Conclusions

The current data reveal evidence that greater aerobic fitness may have benefits for online language processing and aspects of academic achievement in children. Higher fit children have faster neuroelectric responses associated with the appreciation of meaning during reading, and they elicit overall larger N400 amplitude, perhaps due to having richer lexico-semantic networks. The prominent P600 effect in higher fit children closely resembles established patterns of P3b findings, further suggesting that aerobic fitness may exert its influence across a broad network that mediates how attention-demanding, categorization-like processes unfold. Correspondingly, higher fit children also perform better on cognitive tests including those linked to language skills important for school, thus emphasizing the importance of school-based (e.g., Donnelly et al., 2009) and afterschool (e.g., Kamijo et al., 2011) interventions aimed at improving health and fitness in children. These findings add to the emerging literature indicating that fitness may promote better cognitive and brain health, and provide a preliminary understanding of how increased cardiovascular health may be beneficial to language processing.

Conflict of interest

No conflicting financial or personal interests exist.

Funding source

The National Institute of Child Health and Human Development (NICHD) had no involvement in the study design, or in the collection, analysis, and interpretation of data. The NICHD also held no role in the writing or submission of the enclosed manuscript.

Acknowledgment

Support for our research and the preparation of this manuscript were provided by a grant from the National Institute of Child Health and Human Development – United States (NICHD R01 HD055352) to Charles Hillman.

Appendix A. Sentence examples

Congruent – I like to eat *apples* and bananas.
Semantic – I like to eat *doors* and bananas.
Syntactic – I like to *apples* eat and bananas.

Congruent – John can build *houses* with his blocks.
Semantic – John can build *candy* with his blocks.
Syntactic – John can *houses* build with his blocks.

Congruent – The babysitter is washing *dishes* in the sink.
Semantic – The babysitter is washing *sandwiches* in the sink.
Syntactic – The babysitter is *dishes* washing in the sink.

Congruent – At the lake Dylan catches *fish* with Grandpa.
Semantic – At the lake Dylan catches *books* with Grandpa.
Syntactic – At the lake Dylan *fish* catches with Grandpa.

Congruent – It is fun to blow *bubbles* and watch them pop.
Semantic – It is fun to blow *apples* and watch them pop.
Syntactic – It is fun to *bubbles* blow and watch them pop.

Congruent – At school we sing *songs* and dance.
Semantic – At school we sing *cakes* and dance.
Syntactic – At school we *songs* sing and dance.

Congruent – When Grandma visits she brings *candy* for me.
Semantic – When Grandma visits she brings *hands* for me.
Syntactic – When Grandma visits she *candy* brings for me.

Congruent – Mommy is putting *keys* in her purse.
Semantic – Mommy is putting *houses* in her purse.
Syntactic – Mommy is *keys* putting in her purse.

Congruent – George fixes *cars* in his garage.
Semantic – George fixes *bubbles* in his garage.
Syntactic – George *cars* fixes in his garage.

Congruent – The puppy likes to eat *food* from the red dish.
Semantic – The puppy likes to eat *stories* from the red dish.
Syntactic – The puppy likes to *food* eat from the red dish.

References

American College of Sports Medicine (2010). *ACSM's guidelines for exercise testing and prescription* (8th ed.). New York: Lippincott Williams & Wilkins.
Atchley, R. A., Rice, M. L., Betz, S. K., Kwasny, K. M., Sereno, J. A., & Jongman, A. (2006). A comparison of semantic and syntactic event related potentials generated by children and adults. *Brain and Language, 99*, 236–246.
Aydelott, J., Dick, F., & Mills, D. L. (2006). Effects of acoustic distortion and semantic context on event-related potentials to spoken words. *Psychophysiology, 43*, 454–464.
Bar-Or, O. (1983). *Pediatric sports medicine for the practitioner: From physiologic principles to clinical applications*. New York: Springer-Verlag.
Baskin, M. L., Ard, J., Franklin, F., & Allison, D. B. (2005). Prevalence of obesity in the United States. *Obesity Reviews, 6*, 5–7.
Birnbaum, A. S., Lytle, L. A., Murray, D. M., Story, M., Perry, C. L., & Boutelle, K. N. (2002). Survey development for assessing correlates of young adolescents' eating. *American Journal of Health Behavior, 26*, 284–295.
Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research, 1446*, 127–143.
Byrne, J. M., Connolly, J. F., MacLean, S. E., Dooley, J. M., Gordon, K. E., & Beattie, T. L. (1999). Brain activity and language assessment using event-related potentials: Development of a clinical protocol. *Developmental Medicine & Children Neurology, 41*, 740–747.
Byrne, J. M., Dywan, C. A., & Connolly, J. F. (1995). Assessment of children's receptive vocabulary using event-related bran potentials: Development of a clinically valid test. *Child Neuropsychology, 1*, 211–223.
Castelli, D. M., Hillman, C. H., Buck, S. M., & Erwin, H. E. (2007). Physical fitness and academic achievement in 3rd and 5th grade students. *Journal of Sport & Exercise Psychology, 29*, 239–252.
Centers for Disease Control and Prevention (2010). *The association between school-based physical activity, including physical education, and academic performance*. Atlanta, US: Department of Health and Human Services.
Chaddock, L., Erickson, K. I., Prakash, R. S., Kim, J. S., Voss, M. W., VanPatter, M., et al. (2010a). A neuroimaging investigation of the association between aerobic fitness, hippocampal volume and memory performance in preadolescent children. *Brain Research, 1358*, 172–183.
Chaddock, L., Erickson, K. I., Prakash, R. S., VanPatter, M., Voss, M. W., Pontifex, M. B., et al. (2010b). Basal ganglia volume is associated aerobic fitness in preadolescent children. *Developmental Neuroscience, 32*, 249–256.
Chatrian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activity. *American Journal of EEG Technology, 25*, 83–92.
Coch, D., & Holcomb, P. J. (2003). The N400 in beginning readers. *Developmental Psychobiology, 43*, 146–166.
Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33*, 497–505.
Compumedics Neuroscan (2003). *Offline analysis of acquired data (SCAN 4.3) software manual* (4.3 ed., Vol. 2). El Paso, TX.
Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*, 21–58.
Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*(6), 934–945.
Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition, 110*, 293–321.
Donchin, E., & Coles, M. G. H. (1998). Context updating and the P300. *Behavioral and Brain Sciences, 21*, 152–154.

Donnelly, J. E., Greene, J. L., Gibson, C. A., Smith, B. K., Washburn, R. A., Sullivan, D. K., et al. (2009). Physical activity across the curriculum (PAAC): A randomized controlled trial to promote physical activity and diminish overweight and obesity in elementary school children. *Preventive Medicine, 49*, 336–341.

DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (1998). *ADHD Rating Scale-IV: Checklists, norms, and clinical interpretation*. New York: Guilford Press.

Dwyer, T., Sallis, J. F., Blizzard, L., Lazarus, R., & Dean, K. (2001). Relation of academic performance to physical activity and fitness in children. *Pediatric Exercise Science, 13*, 225–237.

Eichenbaum, H., Otto, T., & Cohen, N. J. (1994). Two functional components of the hippocampal memory system. *Behavioral and Brain Sciences, 17*, 449–518.

Eisenmann, J. C., & Malina, R. M. (2002). Secular trend in peak oxygen consumption among United States youth in the 20th century. *American Journal of Human Biology, 14*, 699–706.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Perception and Psychophysics, 16*, 143–149.

Freedson, P. S., & Goodman, T. L. (1993). Measurement of oxygen consumption. In T. W. Rowland (Ed.), *Pediatric laboratory exercise testing: Clinical guidelines* (pp. 91–113). Champaign, IL: Human Kinetics.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *TRENDS in Cognitive Sciences, 6*(2), 78–84.

Friederici, A. D., & Kotz, S. A. (2003). The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage, 20*, S8–S17.

Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience, 16*, 1465–1477.

Friedrich, M., & Friederici, A. D. (2006). Early N400 development and later language acquisition. *Psychophysiology, 43*, 1–12.

Frisch, S., Kotz, S. A., Von Cramon, D. Y., & Friederici, A. D. (2003). Why the P600 is not just a P300: The role of the basal ganglia. *Clinical Neurophysiology, 114*, 336–340.

Grunwald, T., Beck, H., Lehnertz, K., Blümcke, I., Pezer, N., Kurthen, M., et al. (1999). Evidence relating human verbal memory to hippocampal N-methyl-D-aspartate receptors. *Proceedings of the National Academy of Sciences, 96*(21), 12085–12089.

Grunwald, T., Lehnertz, K., Heinze, H. J., Helmstaedter, C., & Elger, C. E. (1998). Verbal novelty detection within the human hippocampus proper. *Proceedings of the National Academy of Sciences, 95*(6), 3193–3197.

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *Journal of Cognitive Neuroscience, 16*(7), 1302–1318.

Healthy Active Living and Obesity (HALO) Research Group (2010). *Pre-participation health screening for children*. Children's Hospital of Eastern Ontario Research Institute.

Henderson, L. M., Baseler, H. A., Clarke, P. J., Watson, S., & Snowling, M. J. (2011). The N400 effect in children: Relationships with comprehension, vocabulary and decoding. *Brain and Language, 117*(2), 88–99.

Hillman, C. H., Buck, S. M., Themanson, J. R., Pontifex, M. B., & Castelli, D. M. (2009). Aerobic fitness and cognitive development: Event-related brain potential and task performance indices of executive control in preadolescent children. *Developmental Psychology, 45*(1), 114–129.

Hillman, C. H., Castelli, D. M., & Buck, S. M. (2005). Aerobic fitness and neurocognitive function in healthy preadolescent children. *Medicine & Science in Sports & Exercise, 37*, 1967–1974.

Hillman, C. H., Erickson, K. I., & Kramer, A. F. (2008). Be smart, exercise your heart: Exercise effects on brain and cognition. *Nature Reviews Neuroscience, 9*, 58–65.

Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology, 8*(2–3), 203–241.

Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience, 14*(6), 938–950.

Institute of Medicine of the National Academies (2013). Educating the student body: Taking physical activity and physical education to school. In H. W. Kohl III, & H. D. Cook (Eds.). Washington DC: The National Academies Press.

Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research, 4*(2), 99–107.

Kamijo, K., Pontifex, M. B., O'Leary, K. C., Scudder, M. R., Wu, C.-T., Castelli, D. M., et al. (2011). The effects of an afterschool physical activity program on working memory in preadolescent children. *Developmental Science, 14*, 1046–1058.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test second edition manual*. Bloomington, MN: NCS Pearson.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*, 627–633.

Kuĉera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23–49.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463–470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205.

Laszlo, S., & Federmeier, K. D. (2007). Better the DVL You Know: Acronyms reveal the contribution of familiarity to single-word reading. *Psychological Science, 18*(2), 122–126.

Laszlo, S., & Federmeier, K. D. (2008). Minding the PS, queues, and PXQs: Uniformity of semantic processing across multiple stimulus types. *Psychophysiology, 45*(3), 458–466.

Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology, 48*(2), 176–186.

Laszlo, S., & Plaut, D. C. (2012). A neurally plausible Parallel Distributed Processing model of event-related potential word reading data. *Brain and Language, 120*(3), 271–281.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)Constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920–933.

Lee, C.-L., & Federmeier, K. D. (2012). In a word: ERPs reveal important lexical variables for visual word processing. In M. Faust (Ed.), *Handbook of the neuropsychology of language* (pp. 184–208). West Sussex, UK: Blackwell.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology, 36*(5), 596–613.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: The MIT Press.

Magnié, M.-N., Bermon, S., Martin, F., Madany-Lounis, M., Suisse, G., Muhammad, W., et al. (2000). P300, N400, aerobic fitness, and maximal aerobic exercise. *Psychophysiology, 37*(03), 369–377.

McCarthy, G., Nobre, A. C., Bentin, S., & Spencer, D. D. (1995). Language-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. *The Journal of Neuroscience, 15*(2), 1080–1089.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience, 7*(7), 703–704.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience, 5*(3), 317–334.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Developmental Neuropsychology, 13*(3), 397–445.

Mills, D. L., Conboy, B. T., & Paton, C. (2005). How learning new words shapes the organization of the infant brain. In L. Namy (Ed.), *Symbol use and symbolic representation* (pp. 123–153). Mahwah, NJ: Lawrence Erlbaum Associates.

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research, 22*(2), 205–220.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*(2), 151–165.

Nobre, A. C., & McCarthy, G. (1995). Language-related field potentials in the anterior-medial temporal lobe: II. Effects of word type and semantic priming. *The Journal of Neuroscience, 15*(2), 1090–1098.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113.

Paffenbarger, R. S., Blair, S. N., & Lee, I.-M. (2001). A history of physical activity, cardiovascular health and longevity: The scientific contributions of Jeremy N Morris, DSc, DPH, FRCP. *International Journal of Epidemiology, 30*(5), 1184–1192.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128–2148.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological Psychology, 41*(2), 103–146.

Pontifex, M. B., Raine, L. B., Johnson, C. R., Chaddock, L., Voss, M. W., Cohen, N. J., et al. (2011). Cardiorespiratory fitness and the flexible modulation of cognitive control in preadolescent children. *Journal of Cognitive Neuroscience, 23*, 1332–1345.

Roberts, C. K., Freed, B., & McCarthy, W. J. (2010). Low aerobic fitness and obesity are associated with lower standardized test scores in children. *The Journal of Pediatrics, 156*(5), 711–718.

Shvartz, E., & Reibold, R. C. (1990). Aerobic fitness norms for males and females aged 6 to 75 years: A review. *Aviation, Space, and Environmental Medicine, 61*(1), 3–11.

Silva-Pereyra, J., Rivera-Gaxiola, M., & Kuhl, P. K. (2005). An event-related brain potential study of sentence comprehension in preschoolers: Semantic and morphosyntactic processing. *Cognitive Brain Research, 23*(2–3), 247–258.

Sitnikova, T., Salisbury, D. F., Kuperberg, G., & Holcomb, P. J. (2002). Electrophysiological insights into language processing in schizophrenia. *Psychophysiology, 39*(06), 851–860.

Sparks, R. L., Patton, J., Ganschow, L., Humbach, N., & Javorsky, J. (2008). Early first-language reading and spelling skills predict later second-language reading and spelling skills. *Journal of Educational Psychology, 100*(1), 162–174.

Tanner, J. M. (1962). *Growth at adolescence: With a general consideration of the effects of hereditary and environmental factors upon growth and maturation from birth to maturity*. Oxford: Blackwell Scientific Publishing.

Taylor, S. J. C., Whincup, P. H., Hindmarsh, P. C., Lampe, F., Odoki, K., & Cook, D. G. (2001). Performance of a new pubertal self-assessment questionnaire: A preliminary study. *Paediatric and Perinatal Epidemiology, 15*(1), 88–94.

Utter, A. C., Robertson, R. J., Nieman, D. C., & Kang, J. (2002). Children's OMNI Scale of Perceived Exertion: Walking/running evaluation. *Medicine & Science in Sports & Exercise, 34*(1), 139–144.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176–190.

Vaynman, S., & Gomez-Pinilla, F. (2005). License to run: Exercise impacts functional plasticity in the intact and injured central nervous system by using neurotrophins. *Neurorehabilitation and Neural Repair, 19*(4), 283–295.

Vaynman, S., & Gomez-Pinilla, F. (2006). Revenge of the "sit": How lifestyle impacts neuronal and cognitive health through molecular systems that interface energy metabolism with neuronal plasticity. *Journal of Neuroscience Research, 84*, 699–715.

Wilkinson, G. S. (1993). *Wide Range Achievement Test 3 administration manual*. Wilmington, DE: Jastak Associates.

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers, 20*(1), 6–10.

Wlotko, E. W., Lee, C.-L., & Federmeier, K. D. (2010). Language of the aging brain: Event-related potential studies of comprehension in older adults. *Language and Linguistics Compass, 4*(8), 623–638.