SUCCESS IN READING… WHAT’S THE MEANING? THE RELATIONSHIP BETWEEN
CHANGES IN CHILDREN’S AEROBIC FITNESS AND LANGUAGE PROCESSING

BY

MARK ROBERT SCUDDER

DISSertation
Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Kinesiology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Charles H. Hillman, Chair
Professor Kara D. Federmeier
Professor Edward McAuley
Associate Professor Steven J. Petruzzello
ABSTRACT

Recent studies have demonstrated that participation in physical activity (PA) programs is a viable means for improving children’s cardiovascular health, including body weight maintenance and increases in aerobic fitness. Additionally, such health outcomes appear to be related to better academic achievement, as well as the underlying cognitive processes governing such performance (i.e., inhibitory control, working memory, etc.). Event-related brain potentials (ERPs) have been instrumental for uncovering further details about the relationship between aerobic fitness and individual aspects of cognitive control; however, very few studies have employed this technique to investigate children’s language processing.

Accordingly, children participated in an after-school PA program over the 9-month academic calendar, while outcome measures were assessed at both pre- and post-test using repeated measures multivariate analysis of variance. In addition to aerobic fitness, demographics, and standardized academic achievement scores, outcome measures included children’s performance on a sentence comprehension task while ERPs were recorded. The N400 and P600 ERP components were of particular interest and provided further information about children’s semantic (i.e., meaning) processing and access to word-related knowledge, as well as their ability to detect syntactic ambiguities and allocate resources towards re-analysis and repair. Secondary hierarchical regression analyses were also conducted to determine the relationship between changes in aerobic fitness and children’s post-test academic performance after controlling for pre-test fitness levels and academic scores, important demographic variables (i.e., age, sex, socioeconomic status [SES], BMI, IQ [intelligence quotient]), and N400/P600 amplitude. Lastly, to replicate prior work, twenty-eight children residing at the lower (≤ 30th percentile) and higher
(≥ 70th percentile) ends of the fitness distribution were matched on age, sex, SES, and IQ, and outcome measures were compared.

Contrary to our hypothesis, children in the intervention group did not exhibit greater increases in aerobic fitness compared to the wait-list control group, yet children in the intervention did display smaller increases in weight and BMI. Given the lack of fitness change, it was not surprising that the intervention group did not experience greater academic gains or increases in sentence performance, nor were there any group ERP differences; however, increases in aerobic fitness were observed among the wave 1 control group (albeit unexpectedly). Interestingly, compared to other participants, greater improvements in academic achievement and sentence performance were witnessed among children in wave 1, with larger increases in academic composite scores occurring primarily in the wave 1 control group versus children in the intervention. Regression analyses also revealed a marginal association between increases in aerobic fitness and greater improvements on standardized tests of reading. This effect was not mediated by the inclusion of children’s N400 and P600 amplitude, which were both independently related to academic performance involving language-based abilities (note: children in wave 1 also had overall larger N400 amplitude compared to other waves). Finally, exploratory comparisons between higher and lower fit children, matched on important demographics, partially replicated findings from an earlier study. Despite no differences in academic performance, higher fit children demonstrated greater sentence task accuracy as well as greater overall N400 amplitude compared to lower fit children (replicating prior findings), yet lower fit children displayed shorter latencies (opposing prior findings).

The current results extend previous work by providing evidence that not only higher aerobic fitness levels, but also increases in children’s fitness may be related to superior
performance on standardized academic achievement measures. Future studies will be necessary to determine if this relationship is indeed selective to reading, or to a broader array of academic domains. Future studies should also continue to design and implement effective PA programs or other strategies that result in the greatest gains for aerobic fitness, especially in terms of targeting and influencing the vast majority of children. Finally, although ERP latency and amplitude changed little over the course of a year and were not associated with changes in fitness, certain component characteristics do appear to be important markers of children’s academic achievement, which may reveal further details if tracked over a longer period of development. Such information would continue to aid researchers in understanding the impact that aerobic fitness and other aspects of cardiovascular health have on children’s cognitive and academic performance, and ensure that programs are optimally designed to deliver the overall greatest possible benefits for children’s health and well-being.
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Hillman and my entire committee for providing years of inspiration, guidance, and support, which have played a vital role in my graduate and professional development. It has been an honor to learn and work besides such notable scholars, and I will continue to draw from this knowledge and experience as I pursue a career in academia.

I would also like to thank the entire Neurocognitive Kinesiology Laboratory (NCK Lab), which has been a 2nd home since my undergraduate years. I was fortunate to work alongside many great colleagues, but I owe a special thanks to Drs. Matt Pontifex, Lauren Raine, and Eric Drollette, who were there from the very beginning. Their mentorship, advice, and friendship were the key ingredients for establishing many successful collaborative efforts, as well as maintaining our busy research agendas.

Lastly, without the love and support of my family and friends I surely would not have made it this far. I am extremely fortunate to be surrounded by such amazing individuals in all aspects of my life.
TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION ............................................................................................................. 1

CHAPTER 2: LITERATURE REVIEW ................................................................................................. 8

CHAPTER 3: METHODS .................................................................................................................. 54

CHAPTER 4: RESULTS ................................................................................................................... 66

CHAPTER 5: DISCUSSION ............................................................................................................... 103

REFERENCES .................................................................................................................................. 116
CHAPTER 1: INTRODUCTION

Schools are widely advocated as a critical environment for expanding children’s opportunities for physical activity (PA) and other healthy behaviors, with the added purpose of negating the effects of an increasingly inactive society (Institutes of Medicine, 2013; Wechsler, Devereaux, Davis, & Collins, 2000). Elevated rates of obesity and the global decline of children’s aerobic fitness levels are among the largest health burdens industrialized nations currently face (Tomkinson, Léger, Olds, & Cazorla, 2003; Wang, McPherson, Marsh, Gortmaker, & Brown, 2011), and such trends involve a substantial proportion of young children and adolescents who endure the concomitant rise of serious health conditions including type 2 diabetes, nonalcoholic fatty liver disease, and metabolic syndrome (Yanovski & Yanovski, 2011). Implementing changes to the school curriculum has been challenging, primarily due to concerns about children’s academic performance if time in the classroom were suddenly allocated to activities outside the confines of a desk (Dwyer, Sallis, Blizzard, Lazarus, & Dean, 2001; Shephard, 1997). However, evidence from a growing body of literature has revealed that more physically active (Coe, Pivarnik, Womack, Reeves, & Malina, 2006; Fisher et al., 2011), more aerobically fit (Castelli, Hillman, Buck, & Erwin, 2007; Chomitz et al., 2009; Eveland-Sayers, Farley, Fuller, Morgan, & Caputo, 2009; Grissom, 2005; Pontifex et al., 2011), and healthy weight children (Kamijo et al., 2012; Li, Dai, Jackson, & Zhang, 2008) exhibit superior cognitive and academic performance across a multitude of tasks. Such findings continue to generate strong interest and innovative approaches for dissociating the relationships between aspects of children’s physiological well-being and scholastic success.

Despite optimistic findings and hopes of promoting children’s health/fitness, as well as academic achievement, school curriculums have remained largely unmodified with regards to the
amount of time allocated to physical education (PE) and other PA opportunities (e.g., recess; McMurrer & Kober, 2007). Perhaps the most obvious reason for the lack of change revolves around the inadequate number of prospective studies and randomized controlled trials supporting a causal relationship between increases in children’s PA/fitness and improvements in academic performance (Biddle & Asare, 2011; Keeley & Fox, 2009). Additionally, the reliance on standardized academic measures has provided limited information about the underlying cognitive processes that are influenced by greater levels of fitness and other important health/demographic variables (e.g., weight status, age, sex, socioeconomic status [SES], etc.). Several studies have now highlighted the importance of cognitive control (also referred to as executive function) for performing deliberate and complex actions such as deciphering difficult tasks, inhibiting prepotent responses/tendencies, and resolving errors to ensure optimal performance (Diamond, Barnett, Thomas, & Munro, 2007; MacDonald, Cohen, Stenger, & Carter, 2000). Multiple cross-sectional investigations across the early childhood years and adolescence have documented the importance of cognitive control (including aspects of inhibitory control, working memory, and cognitive flexibility) for both mathematics and literacy skills (Blair & Razza, 2007; Bull & Scerif, 2001; Gathercole, Pickering, Knight, & Stegmann, 2004), with prospective studies demonstrating that superior cognitive control during the pre-kindergarten years predicts better math and reading performance in primary school (Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Bull, Espy, & Wiebe, 2008). Accordingly, researchers interested in understanding the influence of PA/fitness on children’s academic achievement have expanded their efforts to include additional measures of cognitive processing.

Among the numerous studies that have revealed a beneficial association between greater fitness and academic performance are findings that suggest this relationship is selective to higher
levels of aerobic capacity as compared to other fitness domains (e.g., strength, flexibility, body composition; Castelli et al., 2007; Edwards, Mauch, & Winkelman, 2011; Wittburg, Northrup, & Cottrel, 2009); thus, investigations of children’s cognitive performance have predominately focused on its relation to aerobic fitness. As one might predict, several studies have now documented corroborating results with respect to the positive link between aerobic fitness and aspects of cognitive control (Buck, Hillman, & Castelli, 2008; Mokgothu & Gallagher, 2010; Niederer et al., 2011; Pontifex, Scudder, Drollette, & Hillman, 2012; Scudder et al., 2014b), with additional evidence indicating that more aerobically fit children and adolescents exhibit greater improvements in learning on memory tasks compared to their lesser fit peers (Herting & Nagel, 2012a; Raine et al., 2013). However, the majority of these studies have relied on comparisons between children residing at the highest and lowest fitness levels, thereby disregarding a large and important portion of the population. Longitudinal investigations have demonstrated better cognitive/academic performance and increased aerobic fitness levels among children receiving additional time (Reed, Maslow, Long, & Hughey, 2013) or increased intensity (Ardoy et al., 2013) in PE classes, as well as after-school PA interventions (Davis et al., 2011; Hillman et al., 2014; Kamijo et al., 2011), yet few have verified whether individual changes in aerobic fitness are a contributing factor. As such, reviews of the literature have led to recommendations for further prospective study designs and randomized controlled trials to elucidate the impact of increased aerobic fitness levels on cognitive/academic performance (Smith et al., 2010).

In addition to establishing causality, future studies must rule out other confounding variables that could account for the beneficial influence of aerobic fitness on cognitive control. Poorer cognitive and academic performance have been observed in children as a function of increased adiposity (i.e., body mass index [BMI] or body fat%; Kamijo et al., 2012; Lokken,
Boeka, Austin, Gunstad, & Harmon, 2009), yet other reports have suggested that this relationship disappears once important demographic factors such as SES and parent’s education are considered (Datar, Sturm, & Magnabosco, 2004; Judge & Jahns, 2007). Studies have also demonstrated that fat-free mass, as opposed to the percentage of body fat, is the strongest determinant of VO$_{2\text{max}}$ (the gold-standard measure of aerobic fitness; Goran, Fields, Hunter, Herd, & Weinsier, 2000), which supports the notion that fitness and adiposity should be treated as independent variables when investigating their relationship with cognition (Pontifex et al., 2014). Therefore, regression-based analyses in future work will not only allow researchers to include children across the entire fitness spectrum, but also control for the independent contribution of other important demographic/health factors.

Event-related brain potentials (ERPs) have also demonstrated considerable promise for unveiling information about the association between aerobic fitness and cognitive control, boasting unique advantages compared to expensive neuroimaging methods such as (relatively) low recording costs, minimal invasiveness and subject burden, as well as unmatched temporal resolution that provides a view of underlying cognitive processes with millisecond precision (Kutas & Van Petten, 1994). Fortunately, with the help of ERPs, researchers are well poised to answer the question of how higher aerobic fitness levels and improvements in fitness relate to superior academic performance. Recent work has revealed that the amplitude of certain components is independently associated with better math and reading performance in children, signifying the potential use of ERPs as markers of academic success (Hillman et al., 2012; Khalifian, Stites, & Laszlo, in press). Considering the widespread importance of children’s reading abilities (e.g., phonological awareness, letter knowledge, reading comprehension, etc.) for healthy psychosocial functioning (Boetsch, Green, & Pennington, 1996) and future success in
school (Cunningham & Stanovich, 1997; Leppänen, Aunola, Niemi, & Nurmi, 2008), the
unmatched temporal resolution of ERPs lends itself perfectly for identifying whether specific
reading skills or language processes are amenable to changes in aerobic fitness.

In particular, the N400 component has provided language researchers with a valuable
measure of semantic (i.e., meaning) processing when participants read or listen to individually
presented words and those occurring in a wider sentence/discourse context (Borovsky, Elman, &
Kutas, 2012). Several word characteristics have been shown to increase the amplitude of the
N400, including a larger number of lexical associates (PORK is related to BACON and HAM),
greater orthographic similarity and neighborhood size (PORK/FORK/CORK), as well as higher
frequency of these neighbors (FORK occurs more frequently than TONG; Kučera & Francis,
1967; Laszlo & Federmeier, 2011; Müller, Duñabeitia, & Carreiras, 2010). Hence, N400
amplitude is a useful index of the amount of semantic information arising from lexical access
and the co-activation of distributed features/properties linked to the input (e.g., greater word-
related knowledge; Kutas & Federmeier, 2011). Studies of individual differences in semantic
processing have reported greater N400 amplitude among 1st graders classified as high versus low
reading ability (Coch & Holcomb, 2003), as well as in young adults with high versus low
exposure to print (Laszlo & Sacchi, 2015). Importantly, the amplitude elicited by the critical
word of interest is reduced when it is preceded by predictive/facilitative context, such as a related
or rhyming word (e.g., BACON - PORK; TORQUE - PORK), or a supportive sentence structure:
“Mom cooked barbequed pork/grass for dinner” (the word pork results in lower N400 amplitude
compared to grass; Grossi, Coch, Coffey-Corina, Holcomb, & Neville, 2001; Nobre &
McCarthy, 1994). As such, this reduction in amplitude is referred to as the “N400 effect” and
provides an additional measure of a participant’s ability to utilize supportive context to facilitate semantic processing (Federmeier & Laszlo, 2009; Lau, Phillips, & Poeppel, 2008).

Another ERP component, the P600, has been associated with the analysis of sentence structure (i.e., syntactic processing) with larger P600 amplitude reflecting the detection and re-analysis of syntactic anomalies, including noun/verb disagreements (e.g., “The spoiled child _throw/threw_ the toys on the floor”; Hagoort & Brown, 2000) and phrase structure violations (“The scientist criticized Max’s _of_ proof the theorem”) wherein a sentence becomes difficult to interpret (Hagoort, 2003; Neville, Nicol, Barss, Forster, & Garrett, 1991). It should also be noted that the latency of the N400 and P600 provides additional information regarding the time at which these processes transpire, with several studies documenting shorter latency as a function of increased age (Holcomb, Coffey, Neville, 1992; Friederici, 2005) and greater language proficiency (Moreno & Kutas, 2005; Rossi, Gugler, Friederici, Hahne, 2006).

An initial ERP investigation incorporating children’s aerobic fitness levels found that higher fit individuals displayed superior performance on standardized tests of academic achievement, as well as overall greater N400 amplitude and shorter latency in response to the same target words across sentences (Scudder et al., 2014a). Higher fit children also elicited a large P600 effect when encountering (word-order) violations; an effect that was not witnessed in the lower fit group. The ERP findings suggest higher fit children may have more efficient access to richer word-related knowledge, and an increased ability to detect grammatical violations and engage re-analysis processes. Together, these valuable neuroelectric perspectives hold considerable potential for generating further knowledge regarding the impact of changes in aerobic fitness on aspects of language processing and reading performance.
Purpose and Hypotheses

Accordingly, the current investigation utilized a randomized controlled trial design to examine the impact of a 9-month after-school PA intervention on children’s aerobic fitness levels, as well as related changes in academic achievement, sentence comprehension, and ERP indices of language processing. Outcome measures were compared between intervention and control groups using repeated measures multivariate analysis of variance (RM-MANOVA), while secondary hierarchical regression analyses were used to examine the association between changes in aerobic fitness and academic achievement. It was hypothesized that children in the intervention group would exhibit larger increases in aerobic fitness level, in addition to greater improvements in behavioral performance for academic and sentence task measures. Children in the intervention were also expected to demonstrate increased N400/P600 amplitude and greater reductions in N400 latency, thereby reflecting changes in brain activation associated with superior academic performance. Lastly, to replicate prior work, exploratory analyses were conducted to compare outcome measures between children residing at the lower (≤ 30th percentile) and higher (≥ 70th percentile) ends of the fitness distribution that were matched on age, sex, SES, and IQ. Such findings will help reveal whether higher fitness levels, or increases in aerobic fitness as a result of participation in an after-school PA intervention, are related to greater improvements in children’s academic achievement and modulations of underlying neural activity governing specific aspects of language processing. Such evidence holds significant implications for designing effective PA programs/interventions that can be used in future longitudinal investigations targeting children’s cardiovascular health, cognitive/academic performance, and overall well-being.
CHAPTER 2: LITERATURE REVIEW

Overview

This chapter outlines the declining health trends in children and their potential causes, including how recent changes to school curriculum may be further complicating matters by eliminating physical activity (PA) opportunities in favor of classroom instruction. To demonstrate why this approach might be counterintuitive, findings are reviewed from studies that have documented beneficial relationships between enhanced aspects of children’s physiological health and academic achievement, as well as performance on additional cognitive measures (both cross-sectional and longitudinal evidence). Included in this review are limitations found throughout the literature that are potentially preventing schools from implementing further curriculum changes to circumvent negative health trends and promote better academic achievement. Event-related potentials (ERPs) are introduced as a valuable means of providing a more in depth view of the underlying cognitive processes responsible for children’s academic success, and to determine whether changes in aerobic fitness are related to modulations of specific language/reading skills. Lastly, findings from animal studies and neuroimaging methods in humans are discussed to offer hypotheses about the mechanisms through which aerobic fitness exerts a beneficial influence on brain health and function.

Health Trends in Children

There is considerable evidence indicating that the health of children has been negatively impacted by several adverse societal trends, including increased motorized transport, excess screen time, and the over-consumption of energy-dense foods (Lobstein, Baur, & Uauy, 2004; McDonald, 2007). Such changes have likely played a significant role in the rise of U.S. obesity rates (Finkelstein et al., 2010; Ogden, Carroll, Kit, & Flegal, 2014) and the global decline of
children’s aerobic fitness levels (Olds, Tomkinson, Léger, & Cazorla, 2006; Tomkinson, Léger, Olds, & Cazorla, 2003; Tomkinson, Olds, Kang, & Kim, 2007), leading some experts to suggest that younger generations will be faced with shorter life expectancies compared to their parents for the first time in 200 years (Olshansky et al., 2005). In the U.S., the timing of these events has coincided with the passing of the No Child Left Behind Act of 2001 (NCLB), which is a national initiative focused on improving children’s reading and mathematics performance (as indexed through yearly standardized testing) by increasing the accountability of teachers and schools (Linn, Baker, & Betebenner, 2002). Despite positive intentions, since the initiative’s passing nearly half (44%) of all schools have reported large increases in the time spent on mathematics and reading at the expense of recess and other non-academic subjects, including physical education (PE; McMurrer & Kober, 2007). Coupled with the increased emphasis on academic achievement, commonly-held beliefs that PE is an undervalued/low-demand job (Ennis, 2006) and that PE detracts from academics (Shephard, 1997) have undoubtedly encouraged these changes to school curriculum; however, recent studies have discovered that time spent in PE and recess do not detract from learning or academic performance (Dills, Morgan, & Rotthoff, 2011; Dollman, Boshoff, & Dodd, 2006). In fact, studies have shown the opposite pattern, with data suggesting that a single daily bout of recess is related to better classroom behavior and on-task performance (Barros, Silver, & Stein, 2009; Jarrett et al., 1998). Evidence also indicates that participation in PE and recess during the school day results in children being more physically active after school, thereby providing further protection from sedentary tendencies and associated health risks (Dale, Corbin, & Dale, 2000). Although schools face the difficult challenge of promoting children’s academic achievement, available evidence suggests that the current method of narrowing curricular focus to select academic content is counterproductive,
and that the most effective approaches typically include additional aspects of social, emotional, and physical development (Diamond, 2010).

Children’s Health and Academic Performance

While the prospect of maintaining children’s PA levels and promoting overall wellness in schools is certainly attainable, the current policies set forth by the NCLB act have kept publicly funded schools primarily concerned with children’s academic achievement. Fortunately, due to mandatory standardized academic testing and field estimates of physical fitness recorded by schools, a wealth of data has enabled researchers to reveal a beneficial relationship between these two critical areas. That is, children scoring higher on tests of muscular strength, flexibility, and aerobic fitness demonstrate better academic performance for mathematics, reading, and language arts (Bass, Brown, Laurson, & Coleman, 2013; Colquitt, Langdon, Hires, & Pritchard, 2011; Roberts, Freed, & McCarthy, 2010; Wittberg, Northrup, & Cottrell, 2012), as well as greater school attendance (Blom, Alvarez, Zhang, & Kolbo, 2011); the latter of which is especially important as children must be present to learn. Given the significance of such findings, it is worth mentioning that these results have now been replicated in representative child samples across the globe, on nearly every continent including: Africa (Du Toit, Pienaar, & Truter, 2011), Asia (Desai, Kurpad, Chomitz, & Thomas, 2015; Jacob et al., 2011; Kim et al., 2003; Wang, Wang, & Huang, 2012), Australia (Telford, Cunningham, Telford, & Abharatna, 2012), and Europe (Haapala et al., 2014; Kwak et al., 2009; Torrijos-Nino 2014).

Interestingly, one of the prevailing findings across the literature is that aerobic fitness, as opposed to other fitness domains, exhibits the strongest and most consistent relationship with academic achievement, even after controlling for important demographic variables (de Greeff et al., 2014), including PA levels (Lambourne et al., 2013) and weight status (Rauner, Walters,
Despite requests calling for educators and school officials to incorporate efforts for simultaneous promotion of children’s aerobic fitness and academic performance, the vast proportion of supportive evidence stems from cross-sectional investigations that do not support a causal relationship between increases in aerobic fitness and improvements in academic achievement. Further, the use of children’s overall letter grades, grade point averages, and standardized test scores provides limited information that prevents a mechanistic account regarding the influence of higher aerobic fitness levels on superior academic performance. Therefore, additional research is necessary to determine if changes in aerobic fitness are independently related to modulations in academic achievement after controlling for important health/demographic factors, and whether such changes are associated with differences in brain function/structure or specific cognitive processing abilities.

**Aerobic Fitness and Cognitive Control**

To account for the beneficial relation between higher aerobic fitness and academic achievement, researchers have turned their efforts towards assessing individual aspects of cognitive control (also known as executive function), including: inhibitory control, working memory, and cognitive flexibility, which are responsible for directing goal-oriented behavior and promoting superior academic performance in school (Engle, 2002; Jarrold & Towse, 2006; St Clair-Thompson & Gathercole, 2006). The reasoning for such an approach stems from numerous findings indicating that children who outperform their peers on tests of cognitive control also perform better on academic-based tests (Bull & Scerif, 2001; Borella, Carretti, & Pelegrina, 2010; DeStefano & LeFevre, 2004). It has been further demonstrated that aspects of cognitive control play a significant role in promoting school readiness (more so than intelligence quotient...
and that such performance is predictive of better future math and reading achievement (Diamond & Lee, 2011), as well as greater health and overall well-being as an adult (Moffit et al., 2011). Much like children’s physical maturation, the protracted development of cognitive control extends through adolescence and into young adulthood, as evidenced by age-related changes in brain structure and increased activity of select neural regions (Casey, Galvan, & Hare, 2005; Kwon, Reiss, & Menon, 2002). In particular, studies have shown that children’s cognitive control begins to rely more heavily on portions of the prefrontal and parietal cortex over the course of development, and that this increased involvement is related to better performance across numerous tasks (Bunge & Wright, 2007; Casey, Tottenham, Liston, & Durston, 2005). Given the similar developmental trajectory of both physical and cognitive control abilities, researchers have had ample reason for shifting their focus from academic outcomes to the relationship between aerobic fitness and core aspects of cognition that underlie academic performance. Recent reviews (Fedewa & Soyeon, 2011; Guiney & Machado, 2013) have indicated that the majority of findings (although predominately cross-sectional) suggest a positive association between greater aerobic fitness and superior performance on multiple tasks of cognitive control. Interestingly, recent findings not only support a positive association between these two variables (Khan & Hillman, 2014), but also indicate that this beneficial relationship mediates the direct link between higher fitness and superior academic performance (van der Niet, Hartman, Smith, & Visscher, 2014). Although such findings are influential for elucidating the potential underlying influence of aerobic fitness on children’s academic achievement, once again, the vast majority of supportive evidence stems from behavioral measures and cross-sectional research designs that neither allow for casual inference, nor a direct account for fitness’ influence on brain function and higher-order cognitive processing.
Aerobic Fitness & Cognitive Control: An ERP Perspective

Although changes in behavioral performance on cognitive tasks are of considerable importance and often the primary outcome measure of most interventions, it can be troublesome for researchers to identify causal agents when so many factors are responsible for shaping human behavior. As such, non-invasive neuroimaging and electrophysiological techniques have seen considerable growth given their unique perspectives of the brain’s involvement during cognitive processing. Electroencephalography (EEG) is one technique in particular that has persisted for nearly a century, yet still provides one of the best measures of neural activity by continuously recording voltage differences between electrodes placed on the surface of the scalp (Fabiani, Gratton, & Federmeier, 2007). More recently, researchers have time-locked these recordings to specific cognitive processes that enable the monitoring of an individual’s brain function in preparation for, or in response to, a particular event of interest (hence the name event-related potentials [ERPs]). Despite having a broad array of sensors covering the scalp, ironically, the locations of particular electrodes provide little information about the involvement of specific brain structures due to the variable arrangement of neuron populations and the direction of their electric dipole. However, ERPs gain their advantage by having unrivaled temporal resolution that permits a real-time account of the underlying neural activity elicited by a specific cognitive event (Fabiani et al., 2007; Kutas & Van Petten, 1994).

One component of the ERP waveform that has undergone substantial empirical and theoretical development is the P3, which is thought to reflect attention-driven updating of stimulus representations in working memory (Donchin & Coles, 1998; Polich, 2007). Studies detailing the functional significance of the P3 have demonstrated that its amplitude is sensitive to the availability and amount of attentional resources engaged in task performance, with larger P3
amplitude observed for increased attention (Nieuwenhuis, Aston-Jones, & Cohen, 2005). P3 latency on the other hand serves as an index of stimulus classification/evaluation speed, with shorter latencies reflecting the facilitation of stimulus evaluation (Kutas, McCarthy, & Donchin, 1977; Ridderinkhof & van der Molen, 1995). Early cross-sectional studies assessing the relation of children’s aerobic fitness to ERPs have used this knowledge along with widely implemented cognitive control tasks (e.g., flanker, switch, go-nogo) to help reveal that higher fit children exhibit greater P3 amplitude, shorter P3 latency, and superior behavioral performance (i.e., better accuracy, shorter reaction time [RT]) compared to lower fit children (Hillman, Buck, Themanson, Pontifex, & Castelli, 2009; Hillman, Castelli, & Buck, 2005; Pontifex et al., 2011), suggesting that greater amounts of cardiorespiratory fitness are related to an increased ability to modulate attentional resources and effectively meet task demands. Similar work in adolescents (~14 years old) used a modified go-nogo/flanker task to demonstrate that P3 amplitude was equivalent between higher and lower fitness groups; however, differences were observed for other components of the ERP waveform including decreased N2 amplitude and a larger contingent negative variation (CNV) among higher fit children, suggesting that greater fitness levels may be associated with a reduction in the amount of resources allocated to post-strategic adjustments of conflict monitoring (given equal behavioral performance), as well as increased involvement of anticipatory/attentional processes sub-serving task preparation and stimulus engagement (Stroth, Kubesch, Dieterle, Ruchsow, Heim, Kiefer, 2009). Despite the absence of fitness-related P3 differences in adolescents, which could be due to developmental/maturational factors or nuances in task design/difficulty, ERPs have continued to provide unique and valuable perspectives of the relationship between aerobic fitness and underlying patterns of brain activation linked to core aspects of cognitive processing. Equipped with an effective means of
monitoring differences in cognitive control processes, researchers must now address questions regarding the feasibility of modifying children’s aerobic fitness levels through school-based contexts and/or after-school PA interventions, as well as determine whether changes in fitness are causally related to modulations in cognitive/academic performance.

**Changes in Fitness, Cognitive Control, and Academic Achievement**

Many believe that schools are the ideal environment for introducing programs and interventions designed to improve children’s health (Story, Nanney, & Schwartz, 2009). Fortunately, this logic has generated numerous efforts to modify children’s aerobic fitness and other important health variables (e.g., reductions in body fat%) through modifications to existing PE curriculum or involvement in school-based PA programs, many of which have shown considerable promise (Dobbins, Husson, DeCorby, & LaRocca, 2013). Thus, minor adjustments such as de-emphasizing competition, increasing movement participation (e.g., spending less time changing clothes and standing in lines), and matching the curriculum to children’s skill levels appear to have widespread health benefits, including greater aerobic fitness (Carrel et al., 2005).

There is also evidence to suggest that these approaches are effective for increasing fitness in preschool-age children (~ 4-5 years old; Puder et al., 2011), yet it has been noted that the most effective PA interventions tend to be multidimensional in nature and combined with additional environmental elements such as family and community support (Kriemler et al., 2011). Despite the possibility of introducing these changes to curriculum, there is speculation that schools would be hesitant or slow to implement these methods given the policies set forth by the NCLB act. As such, investigations have also assessed the effectiveness of after-school PA programs for improving aspects of children’s health. For example, Gutin and colleagues (2008) conducted a 3-year after-school intervention focused on accumulating 80-min of daily moderate-to-vigorous PA
in third graders. Compared to the control group, children in the intervention exhibited significant reductions in body fat%, as well as improvements in bone density and aerobic fitness over the 3-year period (Gutin, Yin, Johnson, & Barbeau, 2008). As one might expect, findings throughout the literature involving after-school PA programs closely mirror those of school-based approaches, suggesting that aspects of children’s physiological health are amenable to PA in either setting (Beets, Beighle, Erwin, & Huberty, 2009). Given the potential for improving aerobic fitness through a variety of school-based and after-school contexts, it is of considerable importance to understand the implications that these changes might have for children’s cognitive/academic performance.

Accordingly, numerous PA-based interventions have heightened their focus on academic and cognitive outcomes resulting from children’s participation. Although there is some discrepancy across studies with regard to the specific strategy used to implement PA, the majority of findings indicate a beneficial relation with academic performance whether children participated in an after-school program (Davis et al., 2011), enhanced PE classes (Fisher et al., 2011; Fredericks, Kokot, & Krog, 2006; Sallis et al., 1999; Spitzer & Hollmann, 2013), or classroom-based PA (Erwin, Fedewa, & Ahn, 2012; Hollar et al., 2010; Mahar et al., 2006). Donnelly et al. (2009) demonstrated that second and third grade children who received at least 75 min/week of physically active academic lessons (e.g., multiplication hopscotch) over the course of 3 years not only exhibited smaller increases in body mass index (BMI; i.e., body-weight maintenance), but also displayed greater improvements in mathematics and spelling (Donnelly et al., 2009). Unfortunately, none of these previous studies reported children’s aerobic fitness outcomes, and the majority did not include cognitive control measures to compare differences across intervention/control groups. However, other longitudinal reports have indicated that
children residing at higher aerobic fitness levels, and those who maintain these levels over time, demonstrate superior academic performance (Wittberg, Northrup, & Cottrell, 2012), while additional evidence has suggested that increases in fitness are associated with small improvements in academic achievement (London & Castrechini, 2011). With respect to other cognitive outcomes, studies have shown that children receiving increased time (~ 4-5 days/week) and intensity in PE classes throughout the school year demonstrate greater fitness and larger beneficial changes in perceptual speed (Reed, Maslow, Long, & Hughey, 2013), as well as improved abstract reasoning and spatial ability compared to children receiving standard PE curriculum (~ 1-2 days/week; Ardoy et al., 2013).

Relatively few longitudinal studies have employed ERPs to examine the impact of increased PA/fitness on individual aspects of children’s cognitive control. Findings from the Fitness Improves Thinking in Kids (FITKIds) study revealed that children who participated in a 9-month (after-school) randomized control PA intervention exhibited significant improvements in aerobic fitness, as well as larger accuracy increases on a Sternberg working memory task. At post-test, children in the intervention group also demonstrated greater frontal CNV amplitude during the early part of the waveform following presentation of the to-be-remembered stimulus (i.e., an alphabetical letter). The author’s proposed that the findings reflected an increase in preparatory cognitive processes, such that early CNV differences may indicate a shift to more proactive control strategies characterized by sustained activation and maintenance of goal-relevant information (Kamijo et al., 2011). Stemming from the same intervention, Hillman et al. (2014) assessed children’s inhibitory control and cognitive flexibility using a flanker task and switch task, respectively. Greater positive changes in accuracy performance were witnessed for the intervention group across both tasks, yet this effect was selective for the heterogeneous
(versus homogeneous) switch condition when children had to flexibly shift between rule sets (e.g., shape versus color). Larger increases in P3 amplitude were also observed for the intervention group from pre- to post-test, specifically for the most difficult portions of each task requiring the upregulation of cognitive control (i.e., incongruent flanker trials and heterogeneous switch trials). These intervention findings align closely with previous cross-sectional investigations wherein higher fit children demonstrated selectively greater performance for task conditions placing increased demand on inhibitory control and working memory (Chaddock et al., 2012; Drollette et al., 2016; Pontifex et al., 2011; Voss et al., 2011); however, other studies have reported more general/global fitness benefits using similar cognitive measures (Hillman et al., 2009; Scudder et al., 2014b).

**Understanding the Importance of Individual Changes in Aerobic Fitness**

Although there are several reports highlighting the cognitive benefits associated with improvements in aerobic fitness, another major limitation of previous work is the failure to assess the contribution of individual changes in children’s aerobic fitness while controlling for baseline fitness levels, as opposed to comparing overall group differences. Given the likelihood that individual children may exhibit increased/decreased aerobic fitness levels due to a number of different factors over time, or respond more/less favorably to a PA intervention, regression-based analyses are better-suited for detailing the relationship between changes in aerobic fitness and cognitive processing while controlling for important demographic factors (i.e., socioeconomic status [SES], sex, etc.) that are known to impact cognitive control (Kaufman, 2007; Mezzacappa, 2004). In a large sample of older Swedish male adolescents (18 years old at the time of testing), cross-sectional regression analyses revealed that aerobic fitness (as opposed to muscular strength) was positively related to global IQ, as well as logical and visuospatial
performance scores after controlling for parent’s education level. Interestingly, longitudinal analyses further indicated that participants who demonstrated greater increases in aerobic fitness from 15-18 years old had significantly higher IQ scores, along with superior logical, verbal, and visuospatial scores at 18 years (Åberg et al. 2009). However, note that the interpretation of the change findings was partially based on comparisons between 3 fitness groups (i.e., those who decreased, remained the same, or increased fitness), which were created and identified through regression-based procedures. Lastly, a recent extension of the Donnelly et al. (2009) study incorporated measures of aerobic fitness, inhibitory control (i.e., the flanker task), and working memory (i.e., the n-back task) to further examine the impact of daily physically active academic lessons on second and third grade children’s physical and cognitive health over 3 years (Scudder et al., in press). Cross-sectional regression analyses at both baseline and post-test revealed a well-replicated pattern of results, such that more aerobically fit children demonstrated better performance on both tasks; however, there was further evidence to suggest that as children grew older, the beneficial association between fitness and cognitive performance remained for only the most difficult portions of each task (i.e., flanker incongruent trials and the 2-back working memory condition). As for the longitudinal analyses, no fitness or cognitive differences were observed between intervention and control groups at either baseline or post-test (including equivalent change scores), thus, children were collapsed across groups prior to entering change scores in the regression analyses. After controlling for several important factors, including: grade, sex, mother’s education, BMI, and baseline fitness/cognitive performance, as well as group assignment (intervention versus control), greater increases in aerobic fitness were associated with small but significant improvements in incongruent flanker accuracy and 2-back target discrimination (i.e., increased task difficulty; Scudder et al., in press). Despite uncovering
a relationship between increases in children’s fitness and improvements in cognitive control, the previous study merits the question of what if increases in aerobic fitness were the primary focus of the intervention as opposed to preventing increased BMI? Does the individual pattern of cognitive findings replicate depending on whether changes in fitness are purposeful rather than natural variations over time?

In all, previous studies, while limited by their respective designs, have provided unique evidence demonstrating that higher aerobic fitness levels and increases in fitness are associated with beneficial improvements in academic achievement and cognitive control. Based on each study’s strengths, investigations aiming to advance our understanding of these relationships should strive to include large samples of children spanning a wide distribution of fitness levels, statistically control for important demographic factors known to influence cognitive processing, and conduct analyses that can simultaneously model the effect of a PA intervention alongside individual changes in aerobic fitness (i.e., mixed-model, and hierarchical linear regression).

Further, studies would greatly benefit from the inclusion of standardized academic/cognitive metrics (i.e., achievement/IQ) in addition to ERPs to provide a comprehensive account of the impact of increased aerobic fitness on cognitive processing in children. Such an approach would help elucidate the independent association between modulations in aerobic fitness and cognitive performance, while also accounting for the broad impact that an intervention may have on multiple areas of children’s cognitive/physical development.

**Individual Differences in Language Processing: ERPs as Biomarkers**

As noted previously, there is a large amount of data to suggest that greater aerobic fitness is beneficial for areas of children’s academic achievement, including reading performance, yet there is little evidence to explain how or why such a relationship exists. Given the importance of
early language abilities for promoting continued learning and future academic/economic success (Lonigan, Burgess, & Anthony, 2000; Ritchie & Bates, 2013; Sparks, Patton, Ganschow, Humbach, & Javorsky, 2008), it would be valuable to know whether the cognitive benefits associated with greater aerobic fitness (or changes in fitness) extend across multiple sub-processes involved in reading performance, or demonstrate a select relationship with a particular aspect of language processing that may account for superior behavioral performance on standardized achievement tests. It is well established that general reading ability is comprised of several individual sub-processes working in concert to ensure accurate performance, including: orthographic processing - when readers match printed forms of text with whole-word representations stored in long-term memory (Grainger & Jacobs, 1996), phonological processing - the ability to accurately decode words and syllables into smaller speech/sound segments (Wagner et al., 1997), semantic processing - linking words with their associated meanings (whether presented individually or in a wider sentence/discourse context; Kutas & Federmeier, 2000), and syntactic processing - the analysis of language structure, often with regard to grammaticality, such as word order and subject-verb agreement (Kaan, Harris, Gibson, & Holcomb, 2000). Thus, by determining which of these processes are influenced by greater aerobic fitness levels, or modulations in fitness, researchers would not only gain a better understanding of the underlying factors responsible for superior reading performance in higher fit children, but could potentially use such knowledge to help tailor future PA interventions targeted at groups of children who struggle in these specific language areas.

Fortunately, ERPs have continued to garner praise for their ability to capture specific cognitive events as they unfold in real-time, which has enabled researchers to detect and compare individual differences across a number of distinct processes. As such, ERPs have
become widely regarded as useful biomarkers that can help reveal important information about the neural basis of several clinical disorders (e.g., attention deficit hyperactivity disorder [ADHD] and schizophrenia), in addition to evaluating the effectiveness/outcomes of an intervention (Leppänen, 2013). In fact, there have been suggestions that with enough ERP data collected across multiple children at each age, it would be possible to create large normative datasets that could be used to determine if an individual child’s brain activity matches typically observed patterns associated with normal development, similar to how standardized behavioral measures are used (Yoder, Molfese, Murray, & Key, 2013). Along these same lines, ERP studies have demonstrated that the amplitude and latency of certain components can be used to predict children’s future cognitive/academic performance, even when recorded at very young ages. For example, Molfese (2000) collected auditory ERPs in newborns to test their phonetic discrimination for synthetically produced speech/non-speech sounds, and measured the same individual’s reading performance and IQ at 8 years old using the Wide Range Achievement Test 3 (WRAT3) and the Wechsler Intelligence Scales for Children-3, respectively. Children were separated into 3 groups (control, dyslexic, and poor readers) based on their reading and IQ scores, while amplitude and latency measures of three distinct ERP components were entered as dependent variables in a discriminant function analysis to see if children could be accurately identified as members of their respective group. Results revealed that the discriminant analysis correctly classified ~ 80% of the total sample (well above chance), in addition to 92% of the participants who may have benefitted from early reading intervention (i.e., dyslexic and poor readers; Molfese, 2000). Similar work in young preliterate children has demonstrated the predictive use of ERPs in the visual modality by comparing amplitude differences of the N1 component between individually presented words and pseudowords (i.e., pronounceable
nonwords with no associated meaning such as “wug”). Using multiple regression analysis, Brem et al. (2013) revealed that poorer standardized reading ability in 2nd grade was associated with larger N1 amplitude for words compared to pseudowords (measured with a difference wave) over right occipito-temporal sites during kindergarten. The authors interpreted the smaller N1 amplitude for pseudowords to suggest that poorer readers may have less visual familiarity/expertise with letters and letter strings early in development (Brem et al., 2013).

**ERP Indices of Language Processing**

While previous studies have highlighted the utility of ERP components for predicting children’s early reading performance, many are limited by small sample sizes and the fact that poor versus normal readers are typically grouped according to composite scores taken from multiple standardized reading tests, or individual tests that tap multiple reading sub-processes. Thus, it remains unknown how early childhood differences for specific ERP components, which are often thought to reflect relatively explicit functions, directly relate to individual aspects of reading performance. Such evidence is especially vital for interpreting fitness-related differences in children’s language processing and reading ability using ERPs. For this reason, a number of studies in both children and adults are reviewed below to strengthen the link between specific ERP components and their functional sensitivity with regard to the particular aspects of language they are thought to represent.

**The N400 component.** One of the most widely studied language-related ERP components is the N400 (although it is not entirely language-specific), which is theorized to reflect the brain’s typical response to incoming words that activate stored meaning information in long-term semantic memory (Barber & Kutas, 2007). The N400 was first observed in 1980 by Kutas and Hillyard during a series of studies when participants were asked to read sentences that
ended with semantically incongruent words (“I take coffee with cream and dog”), or appropriate sentence endings that were physically deviant (“I take coffee with cream and SUGAR”). Despite expectations of eliciting a large P3 component to both types of anomalies, a large negative deflection occurred ~ 400 ms following presentation for semantic mismatches only and was largest over parietal-midline sensors. The authors also reported that a higher degree of semantic mismatch resulted in significantly larger N400 amplitude (Kutas & Hillyard, 1980). Since its discovery, researchers have invested considerable time identifying particular aspects of language processing that are associated with modulations of N400 amplitude/latency. Studies have demonstrated that the N400 is present in children as young as 9-12 months old (Friedrich & Friederici, 2010; Parise & Csibra, 2012) and can be reliably elicited using a number of different paradigms involving words, pseudowords, faces, and/or pictures, suggesting that the N400 reflects semantic processing across a host of potentially meaningful stimuli (Lau et al., 2008). The N400 has even been instrumental for revealing that orthographically illegal stimuli attempt lexical access, such as familiar acronyms (i.e., DVD) and unpronounceable letter strings (i.e., DDV; Laszlo & Federmeier, 2007; 2008). Another important detail about the N400 is that it occurs in response to words whether they are presented in isolation, word-pairs, or a larger sentence/discourse context; however, the method of presentation is widely regarded as an important factor influencing overall amplitude (discussed in detail below; Kutas & Federmeier, 2011). Therefore, results across studies/tasks need to be interpreted carefully depending on the particular aspect of the N400 being measured and the paradigm through which the component is derived.

One attribute of the N400 that remains relatively stable across experimental manipulations is the predictable increase in amplitude when greater amounts of lexico-semantic
activation are engendered by an incoming word. It has been widely replicated, both in and out of sentence contexts, that N400 amplitude is larger for known words with greater orthographic neighborhood size (the number of words that share all but one letter in common with the stimulus), higher neighborhood frequency (the average frequency of orthographic neighbors), and more lexical associates (Holcomb, Grainger, & O'Rourke, 2002; Laszlo & Federmeier, 2007, 2011; Midgley, Holcomb, vanHeuven, & Grainger, 2008), which indicates words that share features or are associated with one another tend to become active in parallel (Laszlo & Federmeier, 2009; Laszlo & Plaut, 2012). Recent evidence using linear mixed-effects modeling has further revealed that the influence of orthographic neighborhood size on N400 amplitude remains robust regardless of word position or the amount of accumulated context in a sentence (Payne, Lee, & Federmeier, 2015). With regard to other context-dependent effects, a common pattern of results (matching the original findings of Kutas and Hillyard) has demonstrated that individuals exhibit greater N400 amplitude for words/pictures that are semantically incongruent (versus congruent) with the preceding sentence or discourse context (Van Berkum, 2008), and that this increase in amplitude is inversely proportional to the level of expectedness for the target word, known as a word’s cloze probability (i.e., larger amplitude for more unexpected words; Kutas & Hillyard, 1984). As such, some accounts have proposed that this increase in amplitude, which results in a large difference between congruent and incongruent words (referred to as the N400 effect), reflects greater difficulty integrating words into the preceding context (Brown & Hagoort, 1993; Hagoort, Hald, Bastiaansen, & Petersson, 2004). However, another well-known aspect of the N400 is that the amplitude for a given word is reduced when processing is facilitated through repetition (SOCK - SOCK), or primed by a related/rhyming word (SOCK - SHOE / SOCK - LOCK; Kutas & Federmeier, 2000; Grossi, Coch, Coffey-Corina, Holcomb, &
Neville, 2001). Thus, views about the functional significance of the N400 effect have varied depending on whether the size of the effect is driven by an increase in amplitude from difficulty integrating anomalous words, or a reduction in amplitude from processing words in a facilitative or predictable context.

In support of a facilitative account, researchers have noted that sentences often provide highly constraining and predictive contexts that can pre-activate features of expected words to ease processing (Federmeier, 2007). Lau and colleagues (2009) nicely demonstrated this using magnetoencephalography (MEG) and comparing the N400 effect for semantically related/unrelated word-pairs, as well as sentences with congruent/anomalous endings. The authors reasoned that if integration processes are responsible for the N400 effect, the largest differences should occur during the sentence task when the preceding context is highly structured and integration of an anomalous/unrelated word is more difficult, as opposed to a single preceding word. Further, if integration difficulty is causing this larger difference, the overall N400 amplitude elicited by anomalous sentence endings should be greater than unrelated items occurring 2nd in a word-pair. Interestingly, the results did reveal that the N400 effect was largest during the sentence task; however, it appeared that the size of the effect was primarily driven by a reduction in N400 amplitude for congruent sentence completions. Given that the other trial types had significantly larger amplitude, but did not differ from one another (i.e., anomalous sentence endings did not have the largest amplitude), the authors proposed that the size of the N400 effect is generated by a reduction in the amount of N400 activity a given word elicits when lexico-semantic processing is facilitated (Lau, Almeida, Hines, & Poeppel, 2009). Taken together, evidence spanning the last three decades has revealed that words coming into contact with more information stored in semantic long-term memory result in larger N400 amplitudes,
suggesting that this measure is a reliable index of the richness of the mental lexicon. Relative to this baseline level, the N400 effect can provide an additional measure of how well individuals are able to make use of supportive context during reading, with larger differences reflecting facilitated lexical access (Kutas & Federmeier, 2011).

In contrast to the component’s amplitude, studies have noted that N400 latency remains relatively stable across experimental manipulations, yet it has been known to fluctuate according to two primary factors: age and language expertise. Using both auditory and visual modalities, investigations comparing children and young adults across a wide age range have demonstrated that N400 latency decreases steadily over the course of development (Hahne, Eckstein, & Friederici, 2004; Holcomb et al., 1992; Juottonen, Revonsuo, & Lang, 1996). Although there have been several attempts to generate a similar maturational account for the influence of age on N400 amplitude, such an endeavor has proven difficult due to differences in skull thickness, cortical folding, and other developmental factors that can greatly influence a component’s amplitude (Luck, 2014). To avoid these confounds, ERP recordings in bilinguals, which permit within-participant comparisons, have been helpful for understanding the functional sensitivity of the N400 and how it behaves across languages in more or less experienced users. Investigations adopting this approach have shown that greater language proficiency, both across and within specific languages, is associated with shorter N400 latency (Ardal, Donald, Meuter, Muldrew, & Luce, 1990; Moreno & Kutas, 2005; Weber-Fox & Neville, 1996). Thus, evidence stemming from work conducted among individuals spanning the developmental years, as well as within individuals of a given age, indicate that N400 latency is reduced (i.e., shorter) as a function of greater language expertise and more mature neurocognitive development.
As knowledge about the N400 has continued to grow, a number of studies involving both children and adults have used this information to uncover individual differences in reading/language ability, as well as provide evidence of skill progression and learning when encountering a new (often second) language. For example, findings have demonstrated that young adults with greater exposure to print exhibited larger N400 amplitude for words and ambiguous word items (shaped like the concept they represented) compared to pictures of objects, whereas individuals with lower exposure only displayed increased amplitude for words. The authors interpreted these results to suggest that readers with higher print exposure were able to extract greater orthographic information from ambiguous items in support of more meaningful semantic representations (Laszlo & Sacchi, 2015). Previous work has also compared the N400 in participants who varied on reading comprehension ability, but were matched for IQ and non-word decoding performance. In response to word-pairs, more skilled comprehenders elicited a larger N400 effect over central-midline portions of the scalp that was driven by greater amplitude reductions for related versus unrelated words, particularly when targets were both associatively and categorically related to the probe. Interestingly, this difference was selective to the word-pair task and was not observed when participants viewed pairs of pictures, homophones (BREAK - BRAKE), or orthographically similar non-homophones (GREEN - GREET), providing further indication that semantic processing has an important underlying role in reading comprehension (Landi & Perfetti, 2007). Perfetti et al. (2005) demonstrated similar findings across comprehenders by having participants study rare/difficult words (i.e., CLOWDER) and their associated meanings (i.e., “group of cats”), which were subsequently presented in word pairs together (CLOWDER - CATS) or with the unrelated meaning of another rare word. After just 45 min. of training, skilled comprehenders were more accurate in discriminating related and
unrelated words (i.e., matching rare words to their meanings), and also exhibited a larger N400 effect, suggesting facilitated lexical access for the meanings of words learned just moments earlier (Perfetti, Wlotko, & Hart, 2005). Briefly, it should be noted that the emergence of large N400 effects have also been observed in young adults learning a new language. French learners have been shown to elicit larger N400 amplitude to pseudowords and unrelated words compared to related items, yet this effect was absent in a control group who had never received French instruction. Although this effect became more pronounced over multiple testing sessions as instruction progressed, the researchers further demonstrated that the size of the N400 effect for pseudowords during the first testing session was positively related to the number of instruction hours the participants received (range: 5 - 28 hr.; McLaughlin, Osterhout, & Kim, 2004). Yum et al. (2014) recently reported similar findings among native English speakers (L1) learning Chinese (L2). Fast learners (as determined by larger L2 vocabulary) generated larger overall N400 amplitude for L2 words over frontal portions of the scalp as the number of training sessions progressed, whereas N400 amplitude remained stable for slower learners (Yum, Midgley, Holcomb, & Grainger, 2014).

Previous investigations conducted in adults have been instrumental for understanding the functional significance of the N400 and identifying component characteristics that may be used to differentiate between individuals who exhibit greater improvements in learning or increased proficiency for particular language sub-skills (i.e., comprehension, word recognition, etc.). Therefore, it is promising to see that this work has been extended to younger age ranges with hopes of gathering additional insight about children’s reading development, and using this knowledge to identify less-skilled readers who may benefit from participation in learning opportunities such as reading intervention programs. Similar to findings in adults, Henderson et
al. (2011) revealed that children classified as skilled comprehenders demonstrated a larger N400 effect when judging related and unrelated picture/word pairs. This difference was elicited using spoken words in 8-10 year old children that were divided according to their listening comprehension performance, thereby controlling for word-level reading skills. The authors also observed sizable amplitude differences for earlier components (N100 and N200) that were related to comprehension skill; however, inspection of the waveforms revealed a persistent positive shift for related items beginning around ~ 100 ms, therefore it is unclear which stages of language processing actually contributed to this difference. Smaller overall amplitude (averaged across related and unrelated items) for all three components was also associated with better word recognition and non-word decoding, yet no relationship was observed between N400 amplitude and vocabulary (Henderson, Baseler, Clarke, Watson, & Snowling, 2011). These findings are somewhat discordant with other child studies that have shown better readers and those with greater vocabularies have overall larger N400 amplitude. Coch & Holcomb (2003) separated 1st grade readers according to their composite scores on the Woodcock Reading Mastery Test and compared their ERPs to individual visually presented words, pseudowords, and unpronounceable letter strings while children made a button press any time an animal name appeared. The authors revealed that higher ability readers elicited a clear N400 to all stimulus types, whereas lower ability readers displayed a blunted response resulting in overall smaller amplitude. Although both studies were limited by small sample sizes, the approach used by Coch and Holcomb is more compelling considering they divided participants according to standardized reading scores and compared N400 amplitude to visually presented words, as opposed to listening (i.e., similar task modality). Also, their paradigm had children attend and respond to only animal names during reading, thereby enabling the researchers to compare aspects of language processing
when participants were passively reading critical words/items. This limited the impact of other cognitive/motor processes that could be reflected in the ERP waveform when participants are asked to make judgments and respond on every trial. An elaborate study recently extended these findings to further dissociate the relationships between reading-evoked ERPs and particular aspects of language processing (i.e., phonological awareness, exposure to print, and vocabulary), but perhaps most intriguing, the researchers were interested to see if any of the ERP components were associated with children’s report card performance (pertaining to reading comprehension).

Using a similar protocol to Coch and Holcomb, Khalifian et al. (in press) recorded ERPs in over one hundred children from pre-kindergarten to the age of 7 while they read individually presented words and responded only when their name appeared. However, rather than dividing participants into groups based on performance for a particular reading skill or composite score, the authors performed linear mixed effects regression which allowed them to include and analyze data using each participant’s individual behavioral performance and ERP trials, as opposed to the standard averaging process. Although the authors did not observe a relationship between N400 amplitude and report card scores, larger amplitude was associated with greater vocabulary scores when only real word items were included in the analysis (this finding disappeared when pseudowords and letter strings were included). The only other reading skill that was related to N400 amplitude was phoneme blending, which also had a significant influence on earlier components including the N/P150 and N250. Interestingly, it was the N250 component that was linked to children’s report card scores, which is thought to reflect the point at which phonological processing helps to combine orthographic features into whole word representations (Khalifian et al., in press). In all, research in children, adults, and those learning a second language have uncovered important details concerning modulations of N400 amplitude/latency
and their effectiveness as indices of reading comprehension skill, vocabulary, and general expertise. To grow this body of knowledge, future ERP studies in children should track how these component characteristics change over time within a given individual, which would help solidify current interpretations about their functional sensitivity, as well as determine how/if specific changes reflect increased skill levels or improvements in learning. According to the findings of Khalifian et al. (in press), it may also be beneficial to identify and include multiple prominent ERP components that have been linked to different aspects of language processing, which would allow researchers to continue identifying critical skills that support academic achievement.

**The P600 component.** One particular aspect of language comprehension that the N400 is relatively insensitive to is syntactic processing (Kutas & Federmeier, 2011), which involves the analysis of language structure by combining lexical-level representations to determine how elements in a string of words relate to one another (Fedorenko, Nieto-Castañón, & Kanwisher, 2012). Such processes are critical for accurate comprehension when listening to others or reading sentence passages, and are believed to be partly reflected by the P600 component which has been known to occur primarily over centro-parietal portions of the scalp (Osterhout, McKinnon, Bersick, & Corey, 1996). Early studies documenting the P600 (also referred to as the Syntactic Positive Shift; Hagoort, Brown, & Groothusen, 1993) relied heavily on auditory paradigms that had participants listen to spoken sentences often containing some form of grammatical error, thereby making comprehension difficult. Osterhout & Holcomb (1992) were the first to observe this increased positivity when they had young adults listen to sentences that contained verb subcategorization violations (i.e., “The man hoped to sell the stock” versus “The man persuaded to sell the stock”). The authors noted that the resulting temporary ambiguity created by the
transitive verb “persuaded” renders the participant’s preferred syntactic representation inappropriate (termed the “garden-path” effect), which causes the individual to initiate processes associated with re-analysis and repair. However, since its initial discovery the P600 has been elicited using several different types of syntactic anomalies, including: phrase structure violations - created by inserting incongruous prepositions (“The metal was for refined”; Friederici, Hahne, Mecklinger, 1996) or inverting word order (“Max’s of proof the theorem”; Neville et al., 1991), number/gender agreement (“The hungry guests helped herself to dinner” / “The anxious cowboy prepared herself for the rodeo”; Osterhout & Mobley, 1995), and even for thematic role violations (“For breakfast the eggs would only eat…”), which result in a non-preferred syntactic structure but are grammatically correct (often referred to as “semantic illusions”; Hoeks, Stowe, Doedens, 2004; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003). Given the large number of P600 findings stemming from multiple paradigms and populations, there have been several noted differences in component amplitude, latency, and scalp distribution; however, it has been suggested that this variability may be due to an unknown number of sub-processes that are involved in the formation/representation of the current linguistic input (Brouwer, Fitz, & Hoeks, 2012). Despite the continued debate about the precise functional sensitivity of the P600, many agree that the underlying processes responsible for generating the component are associated with attentionally-demanding and effortful processes of integration, re-analysis, and/or repair (Brouwer, Fitz, & Hoeks, 2012; Friederici, 2002; Kuperberg, 2007).

Just as the N400 was originally believed to be a semantic anomaly “detector”, the P600 was thought to hold a similar role for the recognition of grammatical errors and syntactic ambiguities; however, continued research has argued that the analysis of syntactic structure
occurs much earlier in the waveform, as reflected by frontally-distributed negativities between 100 and 500 ms (ELAN and LAN: [Early] Left Anterior Negativity; Friederici, 2002). While the LAN has been elicited by morphosyntactic errors including subject-verb number agreement ("The elected officials hopes to succeed"; Osterhout & Mobley, 1995), word category errors have been associated with ELANs in multiple studies (Friederici & Kotz, 2003; Neville et al., 1991), yet it is unclear if ELANs are generated entirely by language-based processes or are potentially related to an enhanced sensory response in visual cortex (Dikker, Rabagliati, & Pylkkänen, 2009). In any case, the majority of this evidence has been gathered in adult populations and developmental investigations of the ELAN have suggested that it may not reliably occur to syntactic violations until the teenage years (Hahne et al., 2004). Thus, given the recent number of results demonstrating that initial aspects of syntactic processing are evident prior to semantic access, and that the P600 can be elicited to semantic anomalies devoid of any grammatical error (e.g., thematic role violations and other “semantic illusions”; DeLong, Quante, & Kutas, 2014), there has been an increased emphasis to reevaluate the functional sensitivity of the P600 and its relation to language processing.

Due to the number of findings indicating that the P600 is evoked across a wide range of situations, many theoretical accounts now postulate that the component represents the point at which the products of semantic and syntactic processing interact (Friederici, 2002). Therefore, in certain situations, the outcome of a semantic memory-based analysis may temporarily govern online comprehension, and any conflict detected during the combination of these outputs may lead to re-analysis, as indexed by the P600 (Kuperberg, 2007). However, others have noted that increased P600 amplitude may not necessarily indicate conflict per se, rather, the P600 may simply reflect updating of the current mental/linguistic representation with new information, and
that the size of the P600 effect is largely dependent on the difficulty of task demands (Brouwer et al., 2012). Another important theoretical debate revolves around the findings of Coulson et al. (1998) that highlighted the similar impact of stimulus saliency and probability on the overall amplitude of the P3 and P600 components (i.e., larger amplitude for more salient and less probable stimuli), which led the authors to suggest that the size of the P600 effect may represent an individual’s ability to recognize ungrammatical events and update context accordingly (Coulson, King, & Kutas, 1998).

Interestingly, recent work has highlighted further similarities between the two components by uncovering two dissociable positivities during the P600 time window for highly constraining sentences: a frontal positivity elicited by unexpected but plausible sentence continuations, and a posterior effect evoked by incongruent/anomalous words (DeLong et al., 2014; Van Petten and Luka, 2012). The authors noted that the distinct topographical differences for each component’s maximal amplitude closely matched those of the P3a and P3b, which are associated with the orienting of attention to unexpected or novel stimuli (i.e., frontal P3a), and the updating of context information stored in working memory (i.e., parietal P3b), respectively (Polich, 2007). However, despite the relatedness between the P3 and P600 components, there is evidence in lesion-patients to suggest that the underlying processes responsible for generating each component are supported by relatively distinct brain regions. For example, Frisch et al. (2003) noted that individuals with basal ganglia or temporo-parietal lesions both produce a typical P3 component during Go-NoGo task performance, yet only participants with a temporo-parietal lesion exhibited signs of a syntactic positivity, arguing for the importance of basal ganglia involvement in generating the P600 (Frisch, Kotz, von Cramon, & Friederici, 2003; Kotz, Frisch, von Cramon, & Friederici, 2003). Regardless of the potential differences between
brain regions involved in the production of each component, the common patterns observed for each response across experimental paradigms may suggest a link between these components and the neural/cognitive processes they embody. Future work focusing on such similarities may help reveal further details about the neurocognitive processes underlying aspects of language comprehension and those that mediate information processing in other domains (Osterhout, McKinnon, Bersick, & Corey, 1996). These commonalities are also of critical interest for the current study given the previous fitness findings with respect to greater P3 amplitude witnessed among higher fit children (Hillman et al., 2005, 2009; 2014; Pontifex et al., 2011). Therefore, aerobic fitness levels in children may influence brain areas and/or mechanisms that mediate how attention-demanding, possibly domain-general, processes unfold across multiple types of tasks, including language comprehension.

As with the N400, there has been interest in using the P600 to gather information about children’s language development and the involvement of syntactic processing during comprehension. There is evidence across multiple languages suggesting that children listening to spoken sentences containing syntactic violations begin to elicit a broadly distributed positivity around the age of 3 years. Such an effect has been witnessed in 36-month old English monolinguals that listened to sentences containing auditory verb tense violations (“My Uncle will watching the movie”; Silva-Pereyra, Rivera-Gaxiola, & Kuhl, 2005), and in 31-34 month old German children who encountered phrase structure violations (“The lion in the roars”; Oberecker, Friedrich, & Friederici, 2005). Additional studies attempting to track the developmental progression of syntactic processing have observed reduced P600 amplitude and longer latency among elementary age German children compared to adults (Hahne, Eckstein, & Friederici, 2004), yet a study in English children revealed that adult-like P600 responses were
observed between the ages of 7-8 years old (Atchley et al., 2006). Unfortunately, unlike the N400, the P600 has received relatively little attention with regard to individual differences between children that vary across specific language sub-skill (e.g., reading/listening comprehension, word recognition, etc.); rather, the primary focus of studies has resided on young participants with reading difficulties (e.g., dyslexic children), which have revealed few differences in the P600 time window compared to able readers (Sabisch, Hahne, Glass, von Suchodoletz, & Friederici, 2006). Still, previous research in children and adults spanning multiple languages has established the P600 as a reliable index of processes involved in the detection of syntactic and/or semantic anomalies that impact comprehension and require additional re-analysis/integration with the preceding context. Accordingly, in addition to capturing information about children’s semantic processing using the N400, the P600 can provide further detail regarding the involvement of syntactic processes during accurate reading comprehension. Together these measures will be helpful to further determine which aspects of language processing are related to the observed reading and academic performance differences among higher and lower fit children.

Aerobic Fitness and Language Processing: An ERP Perspective

Despite several findings demonstrating an association between aerobic fitness and cognitive control using ERP components, such as the P3 and CNV, very few studies have adopted the use of language-related components (e.g., the N400 and P600) to compare reading differences among higher and lower fit individuals, or those participating in more/less PA. Previous results have shown that compared to passive learners, adults who were physically active (i.e., riding a cycle ergometer) while learning a second language exhibited not only better vocabulary scores, but also a larger N400 effect following 3 weeks of training sessions with
language-matched word pairs (Schmidt-Kassow, Kulka, Gunter, Rothermich, & Kotz, 2010). However, the sample was underpowered (6 participants in both control and intervention groups) and flaws in the task design, which had individuals perform a lexical decision task to indicate if word pairs matched across languages, seriously limit interpretations of the findings. Given that participants were explicitly paying attention for matching words, it is not surprising that visual inspection of the waveforms revealed both groups elicited a large positivity post-training as their vocabulary/accuracy increased, which may reflect an underlying P3 component to word matches. Therefore, it is difficult to determine if the greater N400 effect in active learners was a function of vocabulary-related differences, or perhaps a larger P3 resulting from the number of correctly matched words between groups.

In another adult study limited by a small sample size, Magnié et al. (2000) found no relationship between aerobic fitness and the N400 component when comparing 10 elite cyclists and sedentary control participants that read sentences with semantically congruous and incongruous endings, yet they did report that overall N400 amplitude for incongruous endings was greater following a single bout of maximal intensity aerobic exercise for all participants. However, rest and exercise sessions were not counterbalanced and separate sentence lists were utilized for each session, which leaves open the possibility that the characteristics of specific lexical items used to evoke the N400 may have impacted the results.

Lastly, a single study in higher and lower fit children (n = 46; 23 per group) compared the N400 and P600 using a sentence reading paradigm that contained semantic or syntactic anomalies 50% of the time, with critical lexical items/words balanced across sentence lists and participants (Scudder et al., 2014a). In addition to better standardized academic achievement scores on the WRAT3, higher fit children demonstrated overall shorter N400 latency and greater
amplitude compared to lower fit children across all sentence types, yet no differences were observed for the size of the N400 effect. Higher fit children also displayed a prominent P600 effect with greater amplitude for syntactic anomalies, however this effect was not witnessed among lower fit children. Together, the N400 latency results revealed that higher fit individuals displayed signs of more mature neurocognitive development (i.e., earlier lexical access), while overall larger amplitude suggested access to greater word-related knowledge. Evidence of a significant P600 effect also indicated that higher fit children were more capable of initiating processes associated with the analysis and/or repair of syntactic errors (Scudder et al., 2014a). It is difficult to draw conclusions about the relationship between aerobic fitness and aspects of language processing based on so little empirical data; however, given that reading ability is undergoing substantial development during the elementary years, this remains a critical period during which aerobic fitness may exert its greatest influence over children’s reading performance. As such, there is a need for additional investigations using ERP indices of language processing to help further elucidate the beneficial relationship between greater aerobic fitness and reading achievement.

Mechanisms Underlying the Fitness/Cognition Relationship

Although not specifically addressed in the current study, questions are often raised concerning the specific changes occurring in the brain that may explain the beneficial relationship between aerobic fitness and cognition. Therefore, a number of exercise/fitness findings are reviewed below from both animal and human studies to provide a comprehensive overview of several mechanisms that may underlie superior cognitive performance: ranging from the healthy growth and development of neurons to increased involvement and communication between key brain structures/regions.
Neurotrophins. Perhaps the most important contribution of animal research has been the ability to examine exercise related changes in the brain at the cellular and molecular levels, which provides a unique opportunity to monitor key molecules that have been known to promote brain health and optimal functioning. The upregulation of proteins such as brain derived neurotrophic factor (BDNF) and insulin-like growth factor-1 (IGF-1) have been closely linked to chronic aerobic exercise participation (Cotman, Berchtold, & Christie, 2007; Foster, Rosenblatt, & Kuljiš, 2011) and support several critical processes in the brain, including (but not limited to): the development and differentiation of neurons, increased synaptic efficiency/transmission via the modulation of neurotransmitter release and vesicle formation, and the overall integration of these components into functional neural circuitry (i.e., neurogenesis; Cotman & Berchtold, 2002; Dishman et al., 2006; Gobeske et al., 2009; Niblock, Brunso-Bechtold, & Riddle, 2000).

Although a number of growth factors display elevated levels following exercise, studies have shown that BDNF and IGF-1 tend to exhibit longer lasting and more robust exercise-related benefits for select brain structures, including the hippocampus, which is a key contributor to learning and memory (Cotman & Berchtold, 2002; Gómez-Pinilla, 2006). In particular, the hippocampus is responsible for relational memory (also known as ‘associative memory’), which involves the binding of converging information to form associations/relationships between objects and events, such as the encoded meanings of individual words and letter strings (Cohen et al., 1999; Henke, Buck, Weber, & Wieser, 1997). Thus, many researchers agree that the improvements in learning and cognitive performance associated with long-term aerobic exercise are at least partially due to greater levels of important neurotrophic factors in critical areas of the brain.
Studies in rodents have shown that habitual time/distance spent on a running wheel demonstrates a positive correlation with levels of BDNF mRNA in the cells of Ammon’s horn areas 1 and 4 (CA1, CA4) of the hippocampus (Neeper, Gómez-Pinilla, Choi, & Cotman, 1995), as well as increased protein levels in CA1-CAl regions (Gomes da Silva et al., 2010). Relative to sedentary animal controls, studies using different learning/memory paradigms have documented greater increases in behavioral performance, elevated BDNF levels, and increased long-term potentiation (LTP) among rodents involved in aerobic exercise (Berchtold, Castello, & Cotman, 2010; Hopkins, Nitecki, & Bucci, 2011; Van Praag, Christie, Sejnowski, & Gage, 1999); mirroring similar effects found in rodents exposed to an enriched living environment (Van Praag, Kempermann, & Gage, 1999). Furthermore, using antagonists to block the effects of BDNF and IGF-1 in select groups of exercising rodents resulted in poorer memory performance analogous to sedentary controls (Cotman et al., 2007; Vaynman, Ying, & Gómez-Pinilla, 2004). In addition to these findings, increased spatial memory performance and expression of BDNF mRNA have been observed in rat pups (~ 1 month post-natal) of mothers who were involved in prolonged aerobic exercise protocols prior to giving birth (Kim, Lee, Kim, Yoo, & Kim, 2007; Parnpiaisil, Jutapakdeegul, Chentanez, & Kotchabhakdi, 2003). Researchers have also proposed that exercise participation (or lack-thereof, see Booth, Chakravarthy, & Spangenburg, 2002) may lead to changes in gene transcripts known to be associated with neuronal health and synaptic plasticity, as demonstrated in rodents (Tong, Shen, Perreau, Balazs, & Cotman, 2001). Human investigations comparing children and adolescents expressing different BDNF gene variations (met- versus val-allele homozygotes) have revealed that individuals carrying the met-allele for BDNF exhibit reductions in hippocampal-to-cortical connectivity among default-mode, executive, and paralimbic networks (Thomason, Yoo, Glover, & Gotlib, 2009). Hopkins and
colleagues (2012) extended these findings by comparing novel object recognition performance and perceived stress among four groups of individuals who participated in a 4-week exercise program, a single bout of exercise on the final testing day, or the combination of both (as well as a non-active control group). Improved recognition performance and reduced stress levels were observed for the combined group, but interestingly, only individuals that were homozygous for the BDNF val-allele demonstrated greater memory performance (as opposed to met carriers; Hopkins, Davis, Vantieghem, Whalen, & Bucci, 2012).

It is also worth noting that aerobic exercise participation and associated increases of neurotrophic factors in rodents have been shown to be protective against a number of brain insults. Following traumatic brain injury, animals allowed access to a running wheel demonstrated comparable BDNF levels and memory performance to sedentary sham rats, both of which were higher than in sedentary concussed rats. However, this effect was only observed when exercise participation was delayed ~ 2 weeks following injury, such that rats partaking in exercise immediately following injury failed to show BDNF upregulation and demonstrated impaired memory performance (Griesbach, Hovda, Molteni, Wu, & Gómez-Pinilla, 2004). Griesbach et al. (2009) replicated these findings using the same delayed exercise protocol, but also revealed that exercising rats who were administered a BDNF antagonist failed to exhibit any improvements in spatial learning, displaying once again the importance of exercise-related increases in BDNF for improved function (Griesbach, Hovda, & Gómez-Pinilla, 2009). Other animal studies have shown protective effects of aerobic exercise for overcoming the deficits of prenatal ethanol exposure, with exercising rats demonstrating similar memory performance and LTP to control animals (Christie et al., 2005). Exercise has also helped rodents recover from neurotoxin insult in the hippocampus, brainstem, and cerebellum, as demonstrated by
comparable spatial memory and motor coordination to controls, yet when an IGF-1 antibody was administered to exercising animals such improvements disappeared (Carro, Trejo, Busiguina, & Torres-Aleman, 2001).

Although a number of previous studies have focused on select proteins (i.e., BDNF and IGF-1), it is worth mentioning that even these factors and their exercise-related benefits have been shown to be mediated by additional variables in the brain. There is some evidence to suggest that exercise-related increases of certain neurotransmitters such as norepinephrine may be responsible for the upregulation of BDNF (Qiang, 2008), whereas other work has highlighted the importance of N-methyl-D-aspartate (NMDA) receptors for mediating increases in BDNF (Kitamura, Mishina, & Sugiyama, 2003) and LTP (Van Praag et al., 1999). Exercise is also effective at reducing bone morphogenetic protein (BMP) signaling, which has been shown to mediate increased hippocampal neurogenesis and cognitive performance (Gobeske et al., 2009).

**Cerebrovascular adaptations.** Another proposed mechanism supporting the beneficial influence of fitness on the brain may be related to adaptations or changes in the underlying cerebrovascular structure to allow greater oxygen availability, as demonstrated in both humans (Brown et al., 2010) and primates. A unique study in aerobically trained monkeys found improved fitness levels and faster learning over a 5 month program, as well as increased vascular volume in the motor cortex, yet this increase disappeared following a 3 month sedentary period (Rhyu et al., 2010). Findings like Rhyu et al. (2010) have been difficult to replicate in humans given the methodology; however, fitness researchers have begun adopting far-less invasive measures that permit further inquiry regarding increased oxygen levels in the brain. Near-infrared spectroscopy (NIRS) is a noninvasive optical imaging technique that enables researchers to measure concentrations of oxy- and deoxyhemoglobin in specific areas of the brain, thanks to
hemoglobin’s unique optical absorption characteristics (Fabiani & Gratton, 2005). For some time researchers have proposed that increased oxygen levels in areas such as the prefrontal cortex (PFC) during moderate intensity aerobic exercise may be related to the acute cognitive benefits observed in humans (Ekkekakis, 2009). Indeed, studies have shown that the rate of increased oxygenation in the PFC is maximal at ~ 42% of the participant’s VO$_{2\text{max}}$ (range: 24.4 - 67.5%; Timinkul, et al., 2008), arguing for the importance of low-moderate intensity exercise; however, relatively few studies have explored this mechanism with respect to individual fitness levels and chronic aerobic exercise participation.

Recent cross-sectional NIRS investigations in both young and older women have revealed that higher fitness levels and reported levels of PA are associated with superior cognitive performance and cerebral oxygenation in areas of the frontal cortex. An investigation in older women reported that compared to lower fit participants, higher fit women demonstrated significantly greater performance on a random number generation task, as well as larger increases in oxygenation within the right dorsolateral PFC; these increases were also found to mediate the beneficial fitness/cognition relationship (Albinet, Mandrick, Bernard, Perrey, & Blain, 2014). Similar results have been observed in groups of younger and older women who were divided into higher and lower fit groups based on their VO$_{2\text{max}}$. Regardless of age, higher fit participants exhibited shorter RTs during the more difficult executive condition of a Stroop task, in addition to greater oxygenation in the right inferior frontal gyrus (Dupuy et al., 2015). Although fitness was not measured, a study of chronic PA levels in young women adds to these findings, which reported that greater activity levels were associated with increased concentrations of oxygenated hemoglobin in anterior portions of the frontal cortex, as well as shorter RT during a task of inhibitory control (Cameron, Lucas, & Machado, 2015). Despite
promising results, further evidence is needed in children and other representative samples (e.g., those including males) to understand if oxygen availability in the brain serves as another important mechanism underlying the fitness-related cognitive benefits observed across the lifespan.

**Changes in brain structure and function.** The final set of evidence that will be reviewed focuses on studies in both adults and children that have uncovered fitness-related differences in brain structure and function with the help of noninvasive magnetic resonance imaging methods (MRI), including functional recordings (fMRI) and diffusion tensor imaging (DTI). The use of such methods in developmental work has been instrumental for detailing the numerous changes that take place in the brain throughout maturation, including significant age-related decreases in cortical gray matter and increases in white matter, despite no overall change in cerebral volume after the age of 5 years (Casey, Giedd, & Thomas, 2000). Accordingly, many researchers acknowledge this time as a period of “rewiring” and synaptic pruning, which has generated questions about the different environmental factors that might influence volumetric changes of specific brain regions and the density/integrity of expanding white matter tracts.

DTI is considered a gold standard measure when it comes to answering the later of these two questions, as it provides information regarding white matter connectivity and microstructure based on the rate and direction of microscopic water diffusion (Boecker, 2011). Sub-measures of DTI commonly include fractional anisotropy (FA) as an indirect measure of myelination and axonal growth, such that higher FA values are associated with greater white matter integrity (Kraus et al., 2007). FA values demonstrate an inverted-U relationship with age, particularly in fronto-temporal connections (Lebel, Walker, Leemans, Phillips, & Beaulieu, 2008), providing further evidence of the protracted development occurring in these regions that eventually
succumbs to age-related decline (Marks et al., 2007). A study in Swedish children and adolescents (8 -18 years) found that elevated FA in the left frontal lobe was associated with superior spatial working memory performance, whereas reading fluency was selectively related to higher FA in the left temporal lobe (Nagy, Westerberg, & Klingberg, 2004). Children with higher FA in the superior longitudinal fasciculus, corona radiata, posterior thalamic radiation, and cerebral peduncle have also exhibited better incongruent accuracy and decreased interference during a modified flanker task (Chaddock-Heyman et al., 2013). Greater FA has even been observed in Broca’s area among young adults who demonstrated better performance on an artificial grammar-classification task involving implicitly learned bi-gram and tri-gram frequencies of consonant strings (Flöel, de Vries, Scholz, Breitenstein, & Johansen-Berg, 2009); although the results were limited by small sample size.

There are relatively few investigations of aerobic fitness and cognition using DTI, yet cross-sectional findings in older adults have revealed significant positive correlations between fitness and FA levels in the corpus callosum, uncinate fasciculus, and cingulum (Johnson, Kim, Clasey, Bailey, & Gold, 2012; Marks et al., 2007). Older adults involved in a one-year aerobic walking intervention exhibited greater increases in fitness compared to a stretching control group (14.6% versus 6.1%), and analyses further demonstrated that larger fitness increases in the intervention group were associated with elevated FA levels among prefrontal and temporal regions, despite a lack of observed change in cognitive performance (Voss et al., 2013). A study in children revealed higher fit participants exhibited greater FA values in the corpus callosum, superior corona radiata, and superior longitudinal fasciculus compared to lower fit individuals, further suggesting greater fitness levels promote the integrity of certain white matter tracts (Chaddock-Heyman et al., 2014). Overweight children involved in an 8-month exercise
intervention also demonstrated improved integrity of the uncinate fasciculus, a tract responsible for communication between temporal and frontal lobes; however, no differences in group aerobic fitness levels were observed at post-test (Schaeffer et al., 2014). One major limitation of these previous studies is the lack of behavioral measures, which makes it difficult to determine if increased white matter integrity serves as a potential mechanism underlying the beneficial association between aerobic fitness and improved cognitive performance. Fortunately, several other imaging investigations using MRI and fMRI have provided additional evidence regarding the influence of aerobic fitness on the volume/function of specific brain structures.

As noted earlier, the protracted growth and development of certain areas in the brain has generated considerable interest for understanding how these progressions unfold, and has helped identify key changes in the brain that are associated with improved cognitive performance. For instance, MRI researchers have revealed that volumetric increases in temporo-parietal white matter (i.e., arcuate fasciculus and superior corona radiata) between kindergarten and 3rd grade are predictive of improvements in children’s reading abilities (e.g., decoding, fluency, and comprehension; Myers et al., 2014). However, with regard to aerobic fitness and brain volume, studies have noted that the observed beneficial relationship between the two variables does not appear to be age-dependent, but is relatively structure-specific (Erickson, Miller, & Roecklein, 2012). Cross-sectional investigations in elementary-age children (Chaddock et al., 2010a), adolescents (Herting & Nagel, 2012a), and older adults (Erickson et al., 2009) have all demonstrated that higher fit individuals exhibit greater hippocampal volume, as well as superior learning and memory performance. Results from these studies also indicated that greater hippocampal volume mediated the positive association between fitness and memory performance (Chaddock et al., 2010a; Erickson et al., 2009). Similar findings have since been replicated in
older adults and patient populations (i.e., schizophrenia) participating in aerobic exercise interventions, which permitted the tracking of changes in fitness and hippocampal volume over time. Compared to a stretching control group, older adults who were involved in a one year aerobic walking intervention demonstrated elevated fitness levels (7.8% versus 1.1% VO$_{2\text{max}}$), improvements in spatial memory performance, and increased volume among anterior portions of the hippocampus, primarily in the dentate gyrus (Erickson et al., 2011; see Pereira et al., 2007 for related findings in mice and humans). Interestingly, greater changes in hippocampal volume among the intervention group were not only related to changes in fitness and memory performance, but to larger increases in serum BDNF as well. Researchers have also witnessed increased hippocampal volume and aerobic fitness levels among healthy adults and patients with schizophrenia that attended a 3-month aerobic exercise program. Compared to non-exercising patients, those from the intervention group exhibited greater improvements in short term memory, and such improvements were related to larger increases in hippocampal volume (Pajonk et al., 2010).

Despite the majority of fitness data focusing on hippocampal volume, it is worth mentioning the cross-sectional findings from Chaddock and colleagues (2010b) that revealed sub-regions of the basal ganglia, including the caudate nucleus, putamen, and the globus pallidus, were larger in higher fit children compared to their lower fit peers. Furthermore, greater volume of the putamen and globus pallidus was associated with better flanker task performance as indicated by greater accuracy for incongruent trials and smaller interference effects; however, no relationships involving ventral striatum (i.e., nucleus accumbens) were observed (Chaddock et al., 2010b). Considering the amount of evidence suggesting aerobic fitness is positively linked to the integrity/volume of the hippocampus and basal ganglia, there may also be observable
fitness differences when comparing performance and language-related ERPs that are supported by these structures. Studies measuring intra-cranial field potentials to anomalous sentence endings and individually presented words have revealed that anterior portions of the medial temporal lobe, including the hippocampus, are involved in generating the N400 (McCarthy, Nobre, Bentin, & Spencer, 1995; Nobre & McCarthy, 1995). Word recognition performance and the size of the N400 are also reduced by ketamine in humans, which serves as an antagonist for NMDA-receptors that are abundant throughout the hippocampus (Grunwald et al., 1999). The amplitude of the ketamine-reduced N400 was also correlated with the density of neurons in the CA1 region of the hippocampus, further demonstrating the importance of this structure for regulating the underlying neural activity sub-serving reading and memory performance.

Similarly, Frisch et al. (2003) have noted the critical involvement of the basal ganglia for eliciting a P600 in lesion patients; therefore, it is intriguing to consider these results alongside the fitness/volume findings of Chaddock et al. (2010b), and the lack of a P600 effect among lower fit children reported by Scudder et al. (2014a). Taken together, the current literature has highlighted important associations between aerobic fitness and the underlying structure of several key regions in the brain; however, future studies are needed to extend this knowledge, including continuing to identify particular aspects of cognition that are amenable to changes in aerobic fitness such as language processing.

 Although the size and integrity of particular brain structures can certainly influence cognitive performance, the effective involvement and communication between these areas is another critical factor that may help explain the cognitive benefits accompanying increased aerobic fitness levels. The final set of evidence that will be summarized stems from investigations using fMRI, which differs from MRI in that the primary outcome or “image” is
derived from changes in blood flow to particular brain structures or regions, thereby reflecting shifts in activation and involvement during cognitive processing (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001). As noted at the beginning of this section, fMRI is an additional imaging method that has been instrumental for providing details about brain activation throughout development, and has helped generate further evidence about the protracted development of the PFC and its involvement in processing different types of information (e.g., verbal, spatial, motor, etc.; Casey et al., 2000).

With regard to aerobic fitness, cross-sectional (Chaddock et al., 2012; Prakash et al., 2011) and longitudinal fMRI studies in both adults and children have revealed that higher fit individuals exhibit greater activation of the PFC and other cortical regions governing attentional control (Weissman, Roberts, Visscher, & Woldorff, 2006), particularly during the most demanding task conditions. Perhaps one of the most convincing pieces of evidence comes from the work of Colcombe et al. (2004), who collected fMRI data from older adults in cross-sectional and longitudinal investigations of aerobic fitness and flanker performance. Cross-sectional findings revealed that greater fitness levels were associated with less behavioral interference (i.e., better performance), as well as increased activation in cortical regions (e.g., middle frontal gyrus, superior frontal gyrus, superior parietal lobe) involved in reorienting attention and maintaining/manipulating contextual information in working memory (Banich et al., 2000). Higher fitness levels were also related to reduced activity in the anterior cingulate cortex (ACC), known for its involvement in conflict monitoring, even after controlling for individual differences in gray matter density (Colcombe et al., 2004). Importantly, these pattern of results were mirrored by longitudinal findings such that older adults who participated in a 6-month aerobic exercise program exhibited not only larger increases in fitness (10.2% versus 2.9%), but
reduced interference scores, larger increases in activation among frontal and parietal cortices, and greater reductions in ACC activity at post-test compared to a stretching/toning control group.

Longitudinal studies in children have revealed similar changes in brain activation and cognitive performance following participation in after-school PA programs. Davis et al. (2011) observed a dose-response increase for standardized measures of planning and math among 3 groups of overweight children who received different durations of daily PA during a 3-4 month intervention (i.e., control, 20 min, and 40 min). Compared to the control group, all children participating in PA also exhibited larger increases in bilateral PFC activation, and reduced activity in bilateral posterior parietal cortex (Davis et al., 2011). However, these results somewhat conflict with the findings of Chaddock-Heyman et al. (2013) who witnessed decreased activation in the right anterior PFC, as well as improvements in incongruent flanker accuracy among children participating in the FITKids intervention (in addition to increased aerobic fitness levels). Interestingly, at post-test children in the intervention showed similar anterior frontal activation patterns and performance on incongruent trials to young adults, yet such changes were not observed in the control group (Chaddock-Heyman et al., 2013). Although the patterns of activation change in critical areas such as the PFC are somewhat discordant across studies tracking participant’s involvement in PA interventions, a number of factors limit interpretations of the current findings including: small sample sizes, the cognitive tasks that were used, the lack of reporting or controlling for the contribution of individual aerobic fitness levels, and the varying demographic/health characteristics between the samples of participants (i.e., age, body composition, etc.). Still, additional cross-sectional evidence in children has indicated aerobic fitness influences activation patterns between specific brain regions and networks, which may reflect the adoption and effective use of cognitive control strategies supporting superior
performance, and could help explain how increased or decreased activation for a particular region relates to improved cognitive processing.

For instance, despite no overall group differences in memory performance, Herting and Nagel (2012b) revealed that higher fit adolescents displayed decreased activation in brain regions that contribute to the default mode network (DMN), whereas lower fit participants demonstrated increased hippocampal and superior frontal gyrus activation during the encoding of later remembered versus forgotten words. Given that the DMN exhibits increased activation during task-irrelevant processes and is typically related to poorer cognitive performance (e.g., longer RT on a selective attention task; Weissman, Roberts, Visscher, & Woldorff, 2006), the authors noted that additional evidence is needed to understand the atypical positive coupling in BOLD response between the hippocampus and DMN during memory performance in lower fit youth (Herting & Nagel, 2012b). The underlying integrity and efficiency of neural networks may also account for the fitness-related performance differences on the flanker task reported by Voss et al. (2011). The results replicated previous findings demonstrating that higher fit children had superior accuracy compared to their lower peers, with larger differences observed for incongruent trials despite lower fit children exhibiting increased activation among brain regions associated with attentional control. In fact, fitness groups showed differential activation patterns such that higher fit children displayed greater activation for congruent versus incongruent trials, whereas the opposite was seen in the lower fit group. The authors proposed that differences in performance and activation patterns may indicate that higher fit children are able to engage proactive cognitive control strategies that allow for flexible adjustments in attention by providing improved maintenance of task demands and stimulus representations (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Voss et al., 2011). In contrast, lower fit children may adopt less-optimal
reactive control strategies that rely on compensatory adjustments in top-down control to avoid errors and resolve conflict (i.e., a “late correction” mechanism; Braver, Paxton, Locke, & Barch, 2009; Pontifex et al., 2011). Differences in control strategies may also account for the fitness differences reported by Chaddock et al. (2012), which revealed that higher fit children maintained incongruent flanker accuracy and exhibited decreased prefrontal/parietal activation as the task block progressed, whereas lower fit participants showed decreased performance and no shifts in activation. While additional fMRI evidence is needed to elucidate the influence of fitness on certain neural networks and cognitive control strategies, the available evidence reviewed herein suggests that higher aerobic fitness levels, as well as greater improvements in fitness, promote optimal function and communication among brain areas supporting cognitive and academic performance. Combined with the numerous other mechanisms linking greater fitness to brain health and function, ranging from increased neurotrophic factors at the cellular level to greater structural volume of important neural regions, there is certainly reason to suggest that changes in fitness are capable of impacting aspects of children’s language processing.
CHAPTER 3: METHODS

Study Design

The current dissertation stemmed from the second iteration of the Fitness Improves Thinking in Kids (FITKids2) randomized controlled trial. The University of Illinois at Urbana-Champaign Institutional Review Board approved the study and all consent/assent documentation. The purpose of the after-school intervention was to provide children with 70 minutes of moderate to vigorous PA per weekday throughout the 9-month school year using a program based on the Community Access to Child Health (CATCH) curriculum, which affords children supervised and instructed lessons to promote age-appropriate physical activities that target improvements in aerobic fitness and other health domains. Elementary-aged children between the ages of 7-9 years were recruited from communities surrounding the University of Illinois at Urbana-Champaign (UIUC) campus and randomized into intervention and wait-list control groups by an individual not involved in testing. Group assignment was also blinded to all research staff involved with data collection. Monetary compensation was provided to all children following the completion of testing, and any child placed in the control group was guaranteed inclusion in the after-school program the following school year.

FITKids2 Intervention

At the end of each school day children were bused to a local recreation center on the UIUC campus. Age-appropriate physical activities for each day were selected from 4-6 instructional activity cards that were part of the CATCH K-5 supplemental materials (Flaghouse, Inc., Hasbrouck Heights, NJ), whereas other activities were adopted from local physical education programs that the children were accustomed to. Activities targeted different aspects of health such as aerobic fitness or muscular strength/endurance, and often involved different
movement concepts and cooperative games. Once a particular activity was completed, children were asked to move to another station where they would interact with a new partner or group during the next activity. In addition to a healthy snack, children received a brief educational component covering topics of fitness, nutrition, or certain health-related tips (e.g., goal-setting, benefits of PA, etc.), which was accompanied by a learning task (e.g., worksheet). Children’s PA/intensity levels were regularly recorded using E600 Polar heart rate monitors (Polar Electro, Finland) and Accusplit Eagle 170 pedometers (San Jose, CA). FITKids intervention leaders were certified physical education teachers that were assisted by college students majoring in kinesiology. Leaders attended an initial three day training workshop to learn how to effectively implement components of the CATCH program while minimizing competitiveness. The entire FITKids staff also attended a one day training workshop that covered several areas of the intervention including: CATCH procedures, educational components, safety and cardiopulmonary resuscitation (CPR) certification, as well as other various FITKids procedures.

**Participants**

During baseline testing, informed assent/consent was provided and explained to the participants and their legal guardians. Guardians confirmed that their child had not received special educational services from their school in connection with learning or attentional disorders, in addition to completing a health history and demographics questionnaire that indicated the child had normal or corrected-to-normal vision, no neurological or attentional difficulties (as further indexed by the ADHD [attention-deficit / hyperactivity disorder] Rating Scale IV; DuPaul, Power, Anastopoulos, & Reid, 1998), and that English was their primary language. Socioeconomic status (SES) was calculated using a trichotomous index based on: (1) highest level of education obtained by the mother and father, (2) number of parents who work
full time, and (3) participation in a free or reduced-price meal program at school (Birnbaum et al., 2002). Guardians also completed the Pre-Participation Health Screening (HALO Research Group, 2010), which is designed to ensure participants are free of pre-existing health conditions that could be exacerbated by physical exercise. Lastly, a modified Tanner Staging System was completed by the guardian and child to ensure all children were pre-pubescent at the time of testing (i.e., a score ≤ 2; Tanner, 1962; Taylor et al., 2001). From a total of 229 participants, children with missing data (n = 58; e.g., did not complete testing at pre- or post-test, missing demographics, etc.), outcome measures exceeding ± 3 SD (n = 5), or those who did not meet aerobic fitness test criteria (n = 2) were excluded from the final analyses, leaving a remaining sample of 164 children.

**Procedure**

The order of testing protocols used during pre- and post-testing were kept identical and divided across two days (with each visit lasting ~ 3 hours). Pre-testing began with participants and guardians completing informed assent and consent on their first visit. While guardians completed registration and demographic forms, participants were administered hand dominance, IQ, and academic achievement tests using the Edinburgh Handedness Inventory (children had to be right handed for inclusion in the study; Oldfield, 1971), Woodcock-Johnson III Test of Cognitive Abilities (WJ-III; Woodcock, McGrew, & Mather, 2001), and Kaufman Test of Academic and Educational Achievement Second Edition (KTEA-2; Kaufman & Kaufman, 2004), respectively. Next, height and weight were measured to calculate body mass index (BMI; kg/m²) before completing a VO₂max test to assess aerobic fitness.

On the second visit, participants were fitted with an electrode cap for neuroelectric measurement and seated in a sound-attenuated room. Participants were given the sentence-task
instructions, made explicitly aware of the two different forms of mistakes, and completed 20 practice sentences (10 congruent, 5 semantic, and 5 syntactic trials) with an identical format to those encountered during the task. Practice sentences were generated from an additional sentence list and were not included in either master list (described in detail below). Participants were afforded the opportunity to ask questions and were given one minute of rest between each block of sentences. After the final block the electrode cap was removed and participants were allowed to wash the electrode gel from their hair.

**Aerobic Fitness**

Children’s maximal oxygen consumption ($VO_{2peak}$) was measured using a computerized indirect calorimetry system (ParvoMedics True Max 2400) during a modified Balke protocol (ACSM, 2010) on a motor-driven treadmill. 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 two minutes using the children’s OMNI Scale (Utter, Robertson, Nieman, & Kang, 2002). Participants were instructed to walk for two minutes before jogging at a constant and comfortable speed, with 2.5% increases in grade every two minutes until volitional exhaustion. Averages for oxygen uptake and respiratory exchange ratio (RER) were calculated every 20 sec. Relative $VO_{2peak}$ was expressed in ml/kg/min (milliliters of oxygen consumed per kilogram of body weight per minute) and evidenced by a minimum of 1 of the following 4 criteria: (1) a plateau in oxygen consumption corresponding to an increase of less than 2 ml/kg/min despite an increase in workload; (2) RER $\geq$ 1.0 (Bar-Or, 1983); (3) a peak HR $\geq$ 185 beats per min (bpm; ACSM, 2010) and a HR plateau (Freedson & Goodman, 1993); or (4) RPE $\geq$ 8 (Utter et al., 2002). Children’s aerobic fitness percentile was calculated using age and sex based normative values (Shvartz & Reibold, 1990).
**Academic Achievement**

Participants were administered composite sections of the KTEA-2, each of which included two subtests that measured different aspects of a particular academic domain. Reading composite scores were derived from performance on word recognition and reading comprehension subtests. For word recognition, participants were asked to identify and pronounce a list of gradually more difficult words, which assessed orthographic familiarity and previous exposure to words. Reading comprehension involved quietly reading short passages/stories, as well as questions related to the passage, and then children provided their answers to the research staff. The reading comprehension portion assessed children’s ability to read and extract the meaning from a set of related sentences (with less of an emphasis on vocabulary level). Next, the fluency composite included subtests of word recognition and nonword decoding, during which participants had to identify and pronounce a list of gradually more difficult real words or pseudowords (i.e., pronounceable nonwords), respectively; however, participants were limited to one minute to correctly read as many words as possible. Nonword decoding relies on phonological decoding skills and the ability to transform printed letters and patterns into sounds for correct pronunciation. Fluency performance provided an additional perspective into children’s reading ability, that being the automaticity of word recognition and phonological decoding, which are critical for overall comprehension and cannot be represented by accuracy measures alone. For the written language composite children completed a spelling portion that assessed the application of phonetic principles and knowledge of letter patterns/sounds by spelling progressively more difficult words, which were pronounced by a trained research staff in a supportive sentence context. Children also completed the written expression subtest, which had children follow along with a set of interesting characters as a researcher narrated a story and
asked participants to write/edit portions of the dialog, captions, or other text. Thus, accurate performance was based on several variables such as correct verb tense, subject-verb agreement, meaningful content, etc. As with the other composites, math was composed of two subtests: concepts and applications, and computation. Concepts and applications required participants to respond orally and apply mathematical principles to real-life situations, which involved skill categories such as number/operation concepts, measurement, and time and money. Computation on the other hand focused strictly on children’s written solutions to sets of increasingly complex math problems that assessed basic operations, fractions, decimals, and more. Lastly, participants completed listening comprehension in which they were asked to actively listen to short passages/stories on a CD recording, and then answer questions related to the meaning of the passage (questions were read by the research staff). The math, written, and listening portions were adopted for the study to determine if aerobic fitness is related to a wider set of academic/language areas, or more select abilities. The overall academic achievement composite score was comprised of reading and math composites, in addition to the written expression and listening comprehension subtests (i.e., six total subtests). The KTEA-2 included two versions of each test, which allowed for repeated assessment.

Sentence Comprehension Task

The eighty sentences in the current paradigm were originally developed for use in the auditory modality with preliterate children, and have been previously used in adults (Aydelott, Dick, & Mills, 2006) and in a cross-sectional study comparing higher and lower aerobically fit children (Scudder et al., 2014a). Sentences were presented using Neuroscan STIM² software (Compumedics, Charlotte, NC) in three different forms: congruent, semantic violation, and word order (i.e., syntactic) violation. Semantic violations were constructed by shuffling target words
across congruent sentences such that the lexical items become semantically anomalous continuations in other contexts.

Congruent sentence A - “You wear shoes on your feet.”

Congruent sentence B - “At school we sing songs and dance.”

For instance, “shoes,” the congruent target word for sentence A above, creates a semantic violation when placed into the target position of sentence B. Syntactic violations were created by reversing the order of the verb and its object noun (congruent: SVO – subject/verb/object; syntactic violation: SOV – subject/object/verb), as demonstrated below:

Congruent - “You wear shoes on your feet.”

Semantic - “At school we sing shoes and dance.”

Syntactic - “You shoes wear on your feet.”

Therefore, the lexical items used as target words remained constant across sentence types.

Characteristics of the target words (number of letters: \( M = 5.1, \ SD = 1.7 \); written frequency: \( M = 99.2, \ SD = 171.9 \), Kučera & 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).

The task contained two master sentence lists that were randomized and counterbalanced across participants. Each list contained all eighty congruent sentences, with the violation versions of each sentence randomized and evenly divided/counterbalanced across the lists for a total of forty semantic and forty syntactic violations in each list (i.e., if master list 1 contained the semantic version of a sentence, master list 2 contained the syntactic version). The order of appearance for each sentence version was also counterbalanced across lists (i.e., congruent 1\textsuperscript{st} -
violation 2nd, or vice versa). That is, if the congruent version of a particular sentence appeared in the first half of master list 1, 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. Master lists were counterbalanced across pre- and post-test. The one hundred and sixty total sentences in each list were divided into eight separate blocks comprised of twenty sentences each (10 congruent, 5 semantic and syntactic each). Sentences ranged from 5 – 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 was presented on the screen asking participants to evaluate what they read (i.e., “Were there any mistakes?”). Participants were asked to respond as quickly and accurately 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 seconds 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 seconds.

**ERP Collection and Reduction**
Electroencephalographic (EEG) activity was recorded from 64 Ag/AgCl sintered 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) while referenced to a midline electrode placed midway between Cz and CPz, with AFz serving as the ground electrode. Electrodes were arranged in an extended montage based on the International 10-20 system (Chatrian, Lettich, & Nelson, 1985) using a Neuroscan Quik-Cap (Compumedics, Charlotte, NC). Bipolar electrooculographic (VEOG/HEOG) activity was recorded using additional electrodes placed above and below the left eye and on the outer canthus of each eye. All impedances were < 10kΩ. Continuous data were filtered with a DC to 70 Hz bandpass filter and a 60 Hz notch filter, amplified 500x using a Neuroscan SynAmps2 amplifier (Compumedics, Charlotte, NC), and sampled at 500 Hz.

Offline, data was re-referenced to averaged mastoids (M1, M2) using Matlab (v.R2012b) and various toolbox plugins from EEGLAB (Delorme and Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2010). A high pass filter of 0.1 Hz (12 dB/oct) was applied prior to conducting independent component analysis (ICA) for identifying stereotypical eye-blink artifact. ICA was subsequently followed by an auto-correlation procedure for rejecting ICA components related to VEOG activity (developed in our lab), which is accomplished by correlating (point-by-point) raw VEOG data with separate ICA activation waveforms (i.e., EEG.icaact matrix produced after running ICA). Components with a correlation coefficient greater than .35 were removed.

Baseline correction was applied using the 100 ms pre-stimulus time window and data were filtered using a 30 Hz (24 dB/oct) low pass filter. Any trials with an artifact exceeding ±75 μV, or those in which a participant answered incorrectly, were automatically rejected. Stimulus-locked epochs were created from -100 to 1000 ms around the target word stimuli, and separate
averages consisted of trials from each sentence type. In accordance with previous findings (Scudder et al., 2014a), a 9-electrode region of interest (Cz, CPz, Pz, C1/2, CP1/2, and P1/2) was used to measure the N400 and P600 components. Mean amplitude and 50% fractional area latency were used to quantify the N400 during the 300-500 ms time window. N400 effect and P600 effect (600-900 ms) amplitude were characterized using the difference of average waveforms between violation (semantic/syntactic) and congruent trials (i.e., violation - congruent trials).

**Statistical Analyses**

All statistical procedures were computed using SPSS v.22 (IBM Corp., Chicago, IL). Children’s demographic/fitness variables (age, SES, pubertal timing, ADHD, IQ, height, weight, BMI, BMI percentile, relative VO$_{2\text{peak}}$, and VO$_{2\text{max}}$ percentile), as well as the different academic composites/subtests, were analyzed using a Time (pre-test, post-test) × Treatment (intervention, control) × Wave (1-4) repeated measures multivariate analysis of variance (RM-MANOVA). To characterize children’s sentence comprehension performance, RT and accuracy were examined using a Time × Trial (congruent, semantic, syntactic) × Treatment × Wave RM-MANOVA. N400 amplitude and latency were also analyzed using this model; however, it should be noted that the Trial factor for the analyses of the N400 and P600 effect only contained two levels (i.e., semantic and syntactic). Post-hoc analyses for significant interactions included independent and paired samples $t$-tests, and Cohen’s $d$ was reported to indicate effect size. Post-hoc comparisons involving Trial or Wave were conducted using one-way ANOVAs with Bonferroni correction. Homogeneity of variances was confirmed using Levene’s test, and in instances of unequal variance, significance was determined using Welch’s $t$-test.
Secondary analyses using hierarchical regression were also conducted to determine if changes in children’s aerobic fitness levels were independently associated with post-test academic achievement, even after controlling for demographics, intervention assignment, wave (using dummy variables), and potential underlying differences in aspects of language processing (i.e., ERPs). Bivariate correlations were analyzed at pre-test between academic, ERP, and demographic/fitness measures (age, sex [female = 0, male = 1], SES, ADHD, IQ, BMI percentile, and VO$_{2\text{max}}$ percentile) to identify significantly related variables that warranted inclusion in the regression model. Pre-test academic measures were regressed onto significant demographic/fitness variables entered in Step 1 (including a group intervention term), wave in Step 2, and significant ERP measures entered in Step 3 to reveal if particular ERP components were independently associated with academic performance. Next, correlations were examined for changes in aerobic fitness and ERP measures, as well as post-test academic performance, to determine if any relation between changes in fitness and academics were potentially mediated by underlying differences in neuroelectric function. Children’s post-test academic achievement was then regressed onto changes in aerobic fitness (Step 3) while controlling for intervention group, pre-test academic performance and aerobic fitness, demographics, and ERPs in Step 1, as well as wave in Step 2. Assumptions of linearity, independent errors, homoscedasticity of the residual terms, and normally distributed errors were plotted, inspected, and verified using studentized residuals. Variance inflation factors were referenced to confirm there were no violations of multicollinearity.

Lastly, exploratory analyses were conducted to replicate prior findings (Scudder et al., 2014a) comparing children residing at the lower ($\leq 30^{\text{th}}$ percentile) and higher ($\geq 70^{\text{th}}$ percentile) ends of the fitness spectrum, matched for important demographic variables (i.e., age, sex, SES,
and IQ). Pre-test demographics and academic performance were compared between lower and higher fit groups using independent t-tests, while sentence behavior was entered into Group (lower fit, higher fit) × Trial (congruent, semantic, syntactic) MANOVAs. Post-hoc comparisons involving Trial were conducted using one-way ANOVAs with Bonferroni correction. Homogeneity of variances was confirmed using Levene’s test, and in instances of unequal variance, significance was determined using Welch’s t-test. Children’s N400 amplitude and latency values were entered into Group (lower fit, higher fit) × Trial (congruent, violation) MANOVAs, whereas the N400 effect and P600 effect for violations trials were compared using independent samples t-tests.
CHAPTER 4: RESULTS

FITKids Intervention

Participants. Mean values for children’s demographics across intervention and control groups are provided in Table 1 (collapsed across wave). Values for any significant differences across waves are reported as (mean ± standard error). Repeated measures multivariate analyses of variance (RM-MANOVAs) revealed no significant effects for socioeconomic status (SES) or parent’s subjective ratings of participant attention-deficit / hyperactivity disorder (ADHD), F’s(1,156) ≤ 2.85, p’s ≥ .09, \( \eta_p^2 \)’s ≤ .018. RM-MANOVAs for age, pubertal timing, height, weight, body mass index (BMI), and BMI percentile demonstrated main effects of Time, F’s(1,156) ≥ 10.29, p’s ≤ .002, \( \eta_p^2 \)’s ≥ .062, reflecting an expected increase in maturation (i.e., higher values at post-test); however, these effects were superseded by Time × Treatment interactions for weight, BMI, and BMI percentile, F’s(1,156) ≥ 4.84, p’s ≤ .029, \( \eta_p^2 \)’s ≥ .030, whereas Time × Wave interactions, F’s(3,156) ≥ 2.92, p’s ≤ .036, \( \eta_p^2 \)’s ≥ .053, were observed for height and weight.

Post-hoc comparisons for the Time × Treatment interactions indicated that intervention and control groups did not differ at pre- or post-test, \( t \)’s(162) ≤ 1.49, p’s ≥ .14, yet the control group exhibited significantly larger increases for weight, BMI, and BMI percentile than those in the intervention, \( t \)’s(162) ≥ 2.29, p’s ≤ .02, \( d \)’s ≥ .36. Paired sample t-tests further revealed that the intervention group did not exhibit a significant increase in BMI percentile, \( t \)(81) = 0.77, p = .44 (i.e., maintenance over time). As for the Time × Wave interactions, post-hoc comparisons for height demonstrated that children in wave 1 (137.46 ± 0.9 cm) were significantly taller than those in wave 4 (133.12 ± 1.0 cm) at pre-test, and that children in wave 4 (4.7 ± 0.3 cm) exhibited a greater increase in height than those in wave 3 (2.89 ± 0.5 cm), F’s(3,160) ≥ 3.18, p’s
≤ .026, d’s ≥ 0.60. No other differences were observed between waves for pretest values or changes in height, p’s ≥ .110, nor were there any post-test height differences, F(3,160) = 1.61, p = .19. For weight, there were no pre- or post-test differences between waves, but despite the significant ANOVA for change in weight, F(3,160) = 2.84, p = .04, post-hoc analyses were non-significant, p’s ≥ .076.

Analyses for intelligence quotient (IQ) revealed a main effect of Time, F(1,156) = 22.55, p < .001, $\eta^2_p = .126$, which was superseded by Time × Treatment and Time × Wave interactions, F’s ≥ 4.39, p’s ≤ .03, $\eta^2_p$’s ≥ .030. Post-hoc comparisons for IQ indicated that children in the control group had higher scores at pre-test than those in the intervention, $t(162) = 2.28$, p < .03, d = 0.36, and only children in the intervention demonstrated a significant increase in standardized scores from pre- to post-test, $t(81) = 4.87$, p < .001, $d = 0.54$ (marginal for control group, p = .059). For the Time × Wave interaction, no pre- or post-test differences were observed between waves, F’s(3,160) ≤ 1.61, p’s ≥ .19; however, wave 1 (5.12 ± 1.4 standard score) and wave 3 (6.57 ± 1.6 standard score) exhibited greater increases in IQ than wave 4 (-0.82 ± 0.8 standard score), p’s ≤ .02, d’s ≥ 0.62. Paired samples t-tests further indicated that both wave 1 and 3 significantly increased from pre- to post-test, $t$’s ≥ 3.60, p’s ≤ .001, and that scores in wave 4 did not change over time, $t(37) = 0.55$, p = .59.

For aerobic fitness, a significant Time × Treatment × Wave interaction was observed for relative VO$_{2peak}$, F(3,156) = 4.01, p = .009, $\eta^2_p = .072$, which nearly reached significance for VO$_{2max}$ percentile, F(3,156) = 2.53, p = .059, $\eta^2_p = .046$. First, to break down this interaction Treatment × Wave univariate ANOVAs were conducted at each time point, which uncovered no between-subject effects/interactions at either time point, F’s ≤ 1.27, p’s ≥ .26, $\eta^2_p$’s ≤ .011. Next,
Time × Treatment RM-MANOVAs conducted for each wave revealed a significant Time × Treatment interaction for wave 1, $F(1,47) = 5.27$, $p = .026, \eta^2_p = .101$, which was marginal for wave 3, $F(1,42) = 3.89$, $p = .055, \eta^2_p = .085$. Further breakdown of these interactions indicated that pre- and post-test relative VO$_{2\text{peak}}$ values were not significantly different across intervention and control groups for wave 1 or 3, $t$’s ≤ 1.41, $p$’s ≥ .16, yet the change in mean fitness values between the two groups was significantly different in wave 1, $t(47) = 2.30$, $p = .026$, $d = 0.70$, and marginal in wave 3, $t(42) = 1.97$, $p = .055$, $d = 0.61$. However, the direction of this relationship was opposite across the two waves, such that the control group (2.46 ± 1.0 ml/kg/min) exhibited a greater increase in fitness compared to the intervention group ( -0.73 ± 0.9 ml/kg/min) in wave 1, whereas the intervention group (1.41 ± 1.4 ml/kg/min) in wave 3 demonstrated a larger fitness increase compared to the control group ( -2.44 ± 1.4 ml/kg/min).

Paired samples t-tests further indicated that the treatment differences in wave 1 were likely due to a significant increase in fitness among the control group, $t(23) = 2.35$, $p = .028$, $d = 0.48$, versus maintained levels in the treatment group, $t(24) = 0.80$, $p = .433$. Similarly, differences in wave 3 may have been primarily due to the marginal decrease in fitness levels among the control group, $t(20) = 1.75$, $p = .096$, compared to the non-significant increase in the treatment group, $t(22) = 1.03$, $p = .312$. Lastly, a Time × Wave RM-MANOVA was conducted separately for both the intervention and control group, which confirmed the findings from the previous interaction breakdowns. A Time × Wave interaction, $F(3,78) = 4.10$, $p = .009, \eta^2_p = .136$, was observed among the control group, yet no significant effects were witnessed for the intervention group, $F$’s ≤ 0.73, $p$’s ≥ .54, $\eta^2_p$’s ≤ .027. Post-hoc analyses among the control group indicated that no fitness differences were observed across waves at pre- or post-test, $F$’s ≤ 1.24, $p$’s ≥ .30, yet wave 1 (2.46 ± 1.0 ml/kg/min) exhibited a larger (and significant; $p = .028$) increase in relative
$\text{VO}_2\text{peak}$ compared to the marginal ($p = .096$) decrease in wave 3 (-2.44 ± 1.4 ml/kg/min), $p = .009$, $d = 0.87$. 
Table 1.

Mean (SE) Values for Participant Demographics across Control and Intervention Groups.

<table>
<thead>
<tr>
<th>Measure (SE)</th>
<th>Control</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
</tr>
<tr>
<td>n</td>
<td>82 (45)</td>
<td>-</td>
</tr>
<tr>
<td>n per Wave (1-4)</td>
<td>24/16/21/21</td>
<td>-</td>
</tr>
<tr>
<td>Race (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>8 (9.8)</td>
<td>-</td>
</tr>
<tr>
<td>Black/African American</td>
<td>12 (14.6)</td>
<td>-</td>
</tr>
<tr>
<td>White/Caucasian</td>
<td>43 (52.4)</td>
<td>-</td>
</tr>
<tr>
<td>Mixed Race or Other</td>
<td>19 (23.2)</td>
<td>-</td>
</tr>
<tr>
<td>Hispanic</td>
<td>6 (7.3)</td>
<td>-</td>
</tr>
<tr>
<td>Age (years)</td>
<td>8.6 (0.1)</td>
<td>9.4 (0.1)</td>
</tr>
<tr>
<td>Pubertal Timing</td>
<td>1.3 (0.1)</td>
<td>1.4 (0.1)</td>
</tr>
<tr>
<td>ADHD</td>
<td>43.1 (3.3)</td>
<td>41.2 (3.3)</td>
</tr>
<tr>
<td>Socioeconomic Status</td>
<td>1.9 (0.1)</td>
<td>1.8 (0.1)</td>
</tr>
<tr>
<td>WJ-BIA (IQ)</td>
<td>111.6 (1.3)</td>
<td>113.6 (1.5)</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>135.5 (0.7)</td>
<td>139.3 (0.7)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>34.4 (1.1)</td>
<td>38.2 (1.2)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>18.5 (0.5)</td>
<td>19.4 (0.5)</td>
</tr>
<tr>
<td>BMI Percentile</td>
<td>64.8 (3.2)</td>
<td>69.0 (3.0)</td>
</tr>
<tr>
<td>Relative VO₂peak (ml/kg/min)</td>
<td>43.0 (0.8)</td>
<td>42.6 (0.8)</td>
</tr>
<tr>
<td>VO₂max Percentile</td>
<td>39.8 (3.4)</td>
<td>37.9 (3.4)</td>
</tr>
</tbody>
</table>

Note. Values sharing a common superscript are not statistically different at \( \alpha = 0.05 \). ADHD – attention-deficit / hyperactivity disorder; WJ-BIA – Woodcock Johnson Brief Intellectual Ability; IQ – intelligence quotient; BMI – body mass index.
**Academic composite scores and subtests.** Given the significant pre-test IQ difference between intervention and control groups, pre-test IQ was entered as a covariate in the academic RM-MANOVAs. Mean values for significant differences between treatment groups and/or waves are reported as (mean ± standard error). Due to several marginally significant Time × Wave interactions for the different academic tests, exploratory post-hoc analyses were conducted to further elucidate the underlying relationships potentially leading to such trends, as noted in the sections below.

**Reading.** The RM-MANOVAs demonstrated a significant Time × Wave interaction for both comprehension and composite scores, F’s(3,155) ≥ 3.60, p’s ≤ .015, η²’s ≥ .065, which was only marginally significant for word recognition, F(3,155) = 2.23, p = .086, η² = .041. Post-hoc comparisons indicated that there were no pre- or post-test differences between waves for any reading measure, F’s(3,160) ≤ 1.74, p’s ≥ .16; however, wave 1 (comprehension: 4.41 ± 1.9; composite: 4.61 ± 1.4 standard score) exhibited a greater increase in reading comprehension and composite scores from pre- to post-test compared to wave 2 (comprehension: -4.70 ± 1.8; composite: -2.94 ± 1.2 standard score), p’s ≤ .013, d’s ≥ 0.76. Paired samples t-tests confirmed that wave 1 demonstrated a significant increase from pre- to post-test, t’s(48) ≥ 2.37, p ≤ .022, d = 0.28, whereas reading scores for wave 2 significantly decreased, t’s(32) ≥ 2.39, p ≤ .023, d = 0.22. Exploratory analyses for word recognition also revealed no differences between waves at pre- or post-test, F’s(3,160) ≤ 1.45, p’s ≥ .23, yet the marginal significance observed for change in word recognition, F(3,160) = 2.30, p = .08, appeared to be primarily driven by a greater improvement in wave 3 (5.80 ± 1.7 standard score), p = .001, d = .31, as compared to maintained performance in wave 2 (-0.39 ± 1.4 standard score), p = .78.
**Fluency.** Significant Time × Wave interactions were observed for both word recognition and nonword decoding fluency subtests, F’s(3,155) ≥ 3.56, p’s ≤ .016, η²_p’s ≥ .064, but was only marginally significant for composite scores, F(3,155) = 2.51, p = .061, η²_p = .046. Post-hoc comparisons indicated that there were no pre- or post-test differences between waves for any fluency measure, F’s(3,160) ≤ 1.02, p’s ≥ .39; however, wave 4 (6.10 ± 1.1 standard score) exhibited a greater increase in performance for word recognition fluency compared to wave 3 (0.54 ± 1.2 standard score), p = .008, d = 0.74, whereas wave 1 (6.39 ± 1.3 standard score) had a larger increase for nonword decoding fluency compared to wave 4 (1.32 ± 1.2 standard score), p = .029, d = 0.61, and nearly wave 2 (1.48 ± 1.7 standard score), p = .054, d = 0.53. Paired samples t-tests confirmed that wave 4 demonstrated a significant increase from pre- to post-test for word recognition fluency, t(37) = 5.68, p < .001, d = 0.42, whereas wave 3 maintained performance, t(43) = 0.43, p = .67. As for nonword decoding fluency, wave 1 significantly increased from pre- to post-test, t(48) = 4.91, p < .001, d = 0.47, yet wave 4, t(37) = 1.11, p = .28, and wave 2 did not exhibit significant changes, t(32) = 0.88, p = .38. Exploratory analyses for fluency composite scores also revealed no differences between waves at pre- or post-test, F’s(3,160) ≤ 0.99, p’s ≥ .40, yet the marginal significance observed for change in composite scores, F(3,160) = 2.24, p = .086, appeared to be primarily driven by a greater improvement in wave 1 (5.75 ± 1.1 standard score), p < .001, d = 0.42, as compared to a smaller but significant increase in wave 3 (2.34 ± 1.0 standard score), p = .017, d = 0.15.

**Written language.** Significant Time × Treatment interactions, F’s(1,155) ≥ 6.13, p’s ≤ .014, η²_p’s ≥ .038, were witnessed for both written composite scores and the written expression subtest. Despite no pre-test differences between intervention and control groups for either measure, t(162) ≤ 1.05, p ≥ .293, post-test composite scores were higher in the control
(composite: 116.00 ± 1.9; expression: 113.80 ± 1.9 standard score) compared to the intervention group (composite: 109.94 ± 1.6; expression: 108.16 ± 1.7 standard score), \( t'(162) \geq 2.19, p's \leq .03, d's \geq 0.34 \). Independent t-tests also indicated that the control group exhibited greater increases for both measures (composite: 6.63 ± 1.1; expression: 8.32 ± 1.6 standard score) compared to the intervention group (composite: 2.96 ± 1.2; expression: 2.15 ± 1.7 standard score), \( t'(162) \geq 2.29, p's \leq .023, d's \geq 0.36 \). Paired samples t-tests confirmed that the control group significantly increased over time for both measures, \( t'(81) \geq 5.26, p's \leq .001, d's \geq 0.51 \), while the intervention group only significantly increased their written composite scores, \( t(81) = 2.50, p = .014, d = 0.21 \) (the smaller increase for expression scores was non-significant, \( p = .208 \)). As for the spelling subtest, analyses returned a marginal Time × Wave interaction, \( F(3,155) = 2.40, p = .070, \eta_p^2 = .044 \). Exploratory analyses revealed no differences between waves at pre- or post-test, \( F'(3,160) \leq 0.84, p's \geq .47 \), yet the marginal significance, \( F(3,160) = 2.21, p = .089 \), observed for change in spelling appeared to be primarily driven by a greater improvement in wave 1 (4.88 ± 1.3 standard score), \( p = .001, d = 0.31 \), compared to maintained performance in wave 2 (0.70 ± 1.2 standard score), \( p = .58 \).

**Math.** The RM-MANOVAs revealed no significant effects for math composite scores, \( F's \leq 1.69, p's \geq .17, \eta_p^2's \leq .032 \). For math computation, a marginal Time × Wave interaction, \( F(3,155) = 2.45, p = .065, \eta_p^2 = .045 \), was observed. Exploratory analyses revealed no differences between waves at pre- or post-test, \( F'(3,160) \leq 0.96, p's \geq .41 \), yet the marginal significance, \( F(3,160) = 2.45, p = .066 \), observed for change in computation scores appeared to be primarily driven by a greater improvement in wave 3 (5.77 ± 2.0 standard score), \( p = .007, d = 0.36 \), compared to maintained performance in wave 1 (-0.26 ± 1.8 standard score), \( p = .88 \), and wave 4 (-0.66 ± 1.8 standard score), \( p = .72 \).
As for math concepts, a main effect of Wave, $F(3,155) = 3.14, p = .027, \eta^2_p = .057$, was superseded by a Time × Treatment × Wave interaction, $F(3,155) = 4.16, p = .007, \eta^2_p = .075$. First, Treatment × Wave univariate ANOVAs were conducted at each time point, which uncovered a marginal effect of Treatment at pre-test, $F(1,155) = 3.14, p = .078, \eta^2_p = .020$, suggesting the control group (109.16 ± 1.6 standard score) had higher baseline performance compared to children in the intervention (105.02 ± 1.6 standard score). At post-test, a marginal Treatment × Wave interaction, $F(3,155) = 2.40, p = .07, \eta^2_p = .044$, was observed. Exploratory post-hoc analyses of this interaction indicated that in wave 4, children in the control group (111.76 ± 4.2 standard score) exhibited greater scores than those in the intervention group (100.88 ± 3.0 standard score), $t(36) = 2.00, p = .053, d = 0.67$. No post-test differences were observed across wave in the control group, $F(3,78) = 0.70, p = .56$; however, among the intervention group, a significant effect of Wave, $F(3,78) = 2.86, p = .042$, indicated wave 3 (115.26 ± 3.4 standard score) had higher scores than wave 4 (100.88 ± 3.0 standard score), $d = 0.98$. Next, Time × Treatment RM-MANOVAs were conducted for each wave, which revealed a significant Time × Treatment interaction for wave 3, $F(1,41) = 12.78, p = .001, \eta^2_p = .238$, while no effects were observed for the other waves, $F$’s ≤ 2.17, $p$’s ≥ .15, $\eta^2$’s ≤ .058. Post-hoc comparisons revealed no significant differences at pre- or post-test between the intervention and control group in wave 3, $t$’s(42) ≤ 1.51, $p$’s ≥ .14, yet the intervention group (8.91 ± 1.8 standard score) exhibited a greater improvement from pre- to post-test, $p < .001, d = 0.55$, as compared to maintained performance ($p = .788$) in the control group (0.52 ± 1.9 standard score), $t(42) = 3.56, p = .001$. Lastly, a Time × Wave RM-MANOVA was conducted separately for both the intervention and control group, which confirmed the findings from the previous interaction.
breakdowns. In the intervention group, a main effect of Wave, $F(1,77) = 3.76, p = .014, \eta^2_p = .128$, was superseded by a Time $\times$ Wave interaction, $F(3,77) = 4.30, p = .007, \eta^2_p = .144$, yet no significant effects were witnessed for the control group, $F's \leq 1.67, p's \geq .18, \eta^2_p's \leq .061$. Post-hoc comparisons in the intervention group revealed no pre-test differences across waves, $F(3,78) = 0.73, p = .53$, yet the ANOVA for post-test performance and change in performance were both significant, $F's(3,78) \geq 2.86, p's \leq .042$. Further breakdown of these effects indicated wave 3 had higher post-test performance ($115.26 \pm 3.5$ standard score), as well as larger increases in performance from pre- to post-test ($8.91 \pm 1.8$ standard score), compared to wave 4 (post: $100.88 \pm 3.0$; change: $0.06 \pm 2.2$ standard score), $p's \leq .028, d's \geq 0.98$. Paired samples t-tests confirmed that the increase in wave 3 was significant, $t(22) = 4.89, p < .001, d = 0.55$, compared to maintained performance in wave 4, $t(16) = 0.03, p = .979$.

Listening comprehension. The RM-MANOVA demonstrated a significant main effect of Wave, $F(3,155) = 4.09, p = .008, \eta^2_p = .073$, which revealed that wave 1 ($111.22 \pm 1.3$ standard score) had overall greater scores compared to wave 2 ($106.62 \pm 1.6$ standard score) and wave 4 ($105.29 \pm 1.5$ standard score), $d's \geq 0.51$, and that wave 3 ($110.62 \pm 1.4$ standard score) also had greater scores than wave 4, $d \geq 0.59$ (and nearly wave 2, $p = .061$). There were also marginal interactions of Time $\times$ Treatment, $F(1,155) = 3.77, p = .054, \eta^2_p = .024$, and Time $\times$ Wave, $F(3,155) = 2.21, p = .089, \eta^2_p = .041$. Exploratory analyses of these interactions indicated that there were no significant listening comprehension differences between control and intervention groups at pre- or post-test, nor was the amount of change different, $t's(162) \leq 1.53, p's \geq .13$. However, paired samples t-tests indicated that the intervention group displayed a marginal decrease in listening performance ($-2.39 \pm 1.2$ standard score), $t(81) = 1.92, p = .058, d = 0.19,$
compared to maintained performance in the control group (0.41 ± 1.3 standard score), \( t(81) = 0.31, p = .76 \). As for the Time × Wave interaction, despite no pre-test differences between waves, \( F(3,160) = 0.01, p = .95 \), a marginal effect was observed for both post-test and change scores, \( F's(3,160) \geq 2.51, p's \leq .06 \), suggesting that wave 1 had higher values compared to wave 4, \( p's \leq .096 \). Paired samples t-tests further indicated that these relationships were likely driven by a significant decrease from pre- to post-test in wave 4 (pre: 109.18 ± 2.0; post: 103.76 ± 2.3 standard score), \( t(37) = 2.70, p = .01, d = 0.40 \), as compared to maintained scores in wave 1 (pre: 109.24 ± 1.9; post: 110.24 ± 1.4 standard score), \( t(48) = 0.58, p = .57 \).

**Overall achievement.** Significant interactions of Time × Treatment and Time × Wave, \( F's \geq 4.12, p's \leq .044, \eta^2_p's \geq .026 \), were witnessed for overall achievement composite scores, which were nearly superseded by a Time × Treatment × Wave interaction, \( F(3,155) = 2.26, p = .084, \eta^2_p = .042 \). Post-hoc comparisons for the Time × Treatment interaction revealed that the control (112.59 ± 1.6 standard score) and intervention (109.35 ± 1.5 standard score) groups did not differ at pre-test, \( t(162) = 1.47, p = .14 \), yet the control group (116.57 ± 1.8 standard score) demonstrated significantly higher scores at post-test compared to the intervention group (110.91 ± 1.6 standard score), \( t(162) = 2.34, p = .02, d = 0.37 \); however, the difference in the amount of change between the two groups was only marginally significant (control: 3.99 ± 1.0; intervention: 1.56 ± 0.9 standard score), \( t(162) = 1.84, p = .068 \). Paired samples t-tests confirmed that the control group significantly increased from pre- to post-test, \( t(81) = 4.20, p < .001, d = 0.26 \), whereas this same relationship was only marginal in the intervention group, \( t(81) = 1.70, p = .093 \) (suggesting maintenance over time). Post-hoc analysis of the Time × Wave interaction indicated that there were no pre- or post-test differences between waves, \( F's(3,160) \leq 0.80, p's \geq .49 \); however, wave 3 (5.80 ± 1.3 standard score) exhibited a greater increase for overall
composite scores from pre- to post-test compared to wave 2 (0.42 ± 1.4 standard score) and wave 4 (-0.10 ± 1.3 standard score), p’s ≤ .032, d’s ≥ 0.64. Paired samples t-tests confirmed that wave 3 demonstrated a significant increase from pre- to post-test, t(43) = 4.41, p < .001, d = 0.38, whereas reading scores for wave 2 and wave 4 maintained over time, t’s ≤ 0.29, p’s ≥ .77. Wave 1 also demonstrated a significant increase (3.88 ± 1.1 standard score) from pre- to post-test, t(48) = 3.39, p = .001, d = 0.25, yet this increase was not significantly different from wave 2 or 4, p’s ≥ .16.

Exploratory analyses of the Time × Treatment × Wave interaction first investigated Treatment × Wave univariate ANOVAs at each time point, which uncovered no significant effects at pre-test, F’s ≤ 2.37, p’s ≥ .13, \( \eta^2_p \)’s ≤ .015, yet a main effect of Treatment, F = 6.36, p = .013, \( \eta^2_p = .039 \), was observed at post-test indicating that the control group (116.77 ± 1.7 standard score) had overall higher scores than children in the intervention (110.56 ± 1.7 standard score). Next, Time × Treatment RM-MANOVAs were conducted for each wave, which revealed a significant Time × Treatment interaction for waves 1 and 2, F’s ≥ 4.27, p’s ≤ .044, \( \eta^2_p \)’s ≥ .085, while no effects were observed for waves 3 and 4, F’s ≤ 1.51, p’s ≥ .23, \( \eta^2_p \)’s ≤ .041. Post-hoc analyses for wave 1 indicated that there were no pre- or post-test differences between control and intervention groups, t’s(47) ≤ 1.17, p’s ≥ .25, yet the control group (6.46 ± 1.8 standard score) exhibited a greater (and significant, p = .002, d = 0.46) increase in performance compared to maintained performance (p = .27) in the intervention group (1.40 ± 1.3 standard score), t(47) = 2.31, p = .025, d = 0.67. For wave 2, there was no pre-test difference between control and intervention groups, t(31) = 1.17, p = .36, yet the control group (119.75 ± 3.3 standard score) had significantly higher scores at post-test compared to the intervention group (109.76 ± 3.0 standard score), t(31) = 2.22, p = .033, d = 0.80. Only a marginal effect was observed between control and
intervention groups for change in performance (control: 3.19 ± 2.1; intervention: -2.18 ± 1.8 standard score), t(31) = 1.92, p = .064; however, paired samples t-tests indicated that neither group exhibited a significant change from pre- to post-test, t’s ≤ 1.48, p’s ≥ .16. Finally, a Time × Wave RM-MANOVA was conducted separately for both the intervention and control group. In the intervention group, a Time × Wave interaction, F(3,77) = 6.76, p < .001, η² = .209, was observed, yet no significant effects were witnessed for the control group, F’s ≤ 1.50, p’s ≥ .22, η²’s ≤ .055. Post-hoc comparisons in the intervention group revealed no pre- or post-test differences across waves, F’s(3,78) ≤ 0.73, p’s ≥ .36, yet the ANOVA for change in performance was significant, F(3,78) = 6.85, p < .001, suggesting wave 3 (7.13 ± 1.7 standard score) had larger increases in academic performance compared to wave 2 (-2.18 ± 1.8 standard score) and wave 4 (-2.00 ± 2.0 standard score), p’s ≤ .002, d’s = 1.13, whereas the same relationship with wave 1 (1.40 ± 1.2 standard score) was marginal, p = .061. Paired t-tests confirmed that wave 3 significantly increased scores from pre- to post-test, t(22) = 4.09, p < .001, d = 0.49, yet the other three waves only maintained performance, t’s ≤ 1.21, p’s ≥ .24.

Sentence task behavior and ERPs.

Participants. In the following analyses children were excluded if they did not complete ERP measurement or have enough usable trials (≥ 15 trials) to compute an average waveform for each of the three trial types at pre- or post-test (n = 41), in addition to sentence accuracy below 50% for any trial type at either time point (n = 4). Mean values for children’s demographics across intervention and control groups are provided in Table 2 (collapsed across wave). Values for any significant differences across waves are reported directly as (mean ± standard error).

---

1 Baseline demographics were compared between included/excluded individuals, which revealed that excluded individuals were slightly younger (8.5 vs 8.7 years, p = .03) and had higher subjective ADHD ratings (50.5 vs. 39.6, p = .03) as reported by their parent/guardian. No other differences were observed (p’s ≥ .12).
RM-MANOVAs revealed no significant effects for parent’s subjective ratings of participant ADHD, $F's \leq 2.10, p's \geq .15, \eta^2_p's \leq .019$, yet for SES there was a significant main effect of Treatment, $F(1, 111) = 4.33, p = .040, \eta^2_p = .038$, indicating that children in the treatment group resided in higher SES households than children in the control group. RM-MANOVAs for age, pubertal timing, weight, BMI, and BMI percentile demonstrated main effects of Time, $F's(1, 111) \geq 9.20, p's \leq .003, \eta^2_p's \geq .077$, reflecting an expected increase in maturation (i.e., higher values at post-test); however, for height and IQ, main effects of Time were superseded by Time x Wave interactions, $F's(3, 111) \geq 2.96, p's \leq .035, \eta^2_p's \geq .075$. Post-hoc comparisons revealed no pre- or post-test differences across waves for height or IQ, $F's(3, 111) \leq 2.14, p's \geq .10$, yet the ANOVA for change in both measures was significant, $F's(3, 115) \geq 2.71, p's \leq .048$, suggesting wave 4 (4.69 ± 0.3 cm) demonstrated larger increases in height compared to wave 3 (2.38 ± 0.7 cm), whereas wave 3 had larger increases in IQ (7.83 ± 1.7 standard score) than wave 4 (0.11 ± 1.5 standard score), $p's \leq .041, d's \geq 0.77$.

For aerobic fitness, a significant Time x Treatment x Wave interaction was observed for relative VO$_{2\text{peak}}$, $F(3, 111) = 3.86, p = .011, \eta^2_p = .094$, yet no significant effects were witnessed for VO$_{2\text{max}}$ percentile, $F's \leq 1.67, p's \geq .18, \eta^2_p's \leq .043$. To break down this interaction Treatment x Wave univariate ANOVAs at each time point were conducted first, which uncovered no between-subject effects or interactions at either time point, $F's \leq 1.82, p's \geq .15, \eta^2_p's \leq .047$. Next, Time x Treatment RM-MANOVAs were conducted for each wave and revealed a significant Time x Treatment interaction for wave 3, $F(1, 27) = 4.66, p = .040, \eta^2_p = .147$, which was marginal in wave 1, $F(1, 31) = 3.34, p = .077, \eta^2_p = .097$. Further breakdown of
these interactions indicated that pre- and post-test relative VO\textsubscript{2peak} values were not significantly different across intervention and control groups for either wave 1 or 3, \(t\)'s \leq 1.72, \(p\)'s \geq .10, yet the change in mean fitness values between intervention and control groups was significantly different in wave 3, \(t(27) = 2.16, p = .04, d = 0.85\), and was marginally significant in wave 1, \(t(31) = 1.83, p = .077, d = 0.66\). The direction of this relationship was opposite across the two waves, such that the intervention group (1.99 \pm 1.3 ml/kg/min) exhibited a greater increase in fitness compared to the control group (-3.29 \pm 2.3 ml/kg/min) in wave 3, whereas the control group (2.24 \pm 1.3 ml/kg/min) demonstrated a larger fitness increase in wave 1 compared to the intervention group (-1.09 \pm 1.3 ml/kg/min); however, paired samples t-tests indicated that neither group in waves 1 or 3 demonstrated a significant change from pre- to post-test \(t\)'s \leq 1.70, \(p\)'s \geq .11. Finally, a Time \times Wave RM-MANOVA was conducted separately for both the intervention and control group. A Time \times Wave interaction, \(F(3,51) = 2.57, p = .065, \eta_p^2 = .131\), was marginal among the control group, but no significant effects were witnessed for the intervention group, \(F\)'s \leq 1.24, \(p\)'s \geq .30, \(\eta_p^2\)'s \leq .059. Exploratory post-hoc analyses among the control group indicated that no fitness differences were observed across waves at pre- or post-test, \(F\)'s \leq 1.03, \(p\)'s \geq .38, although wave 1 (2.24 \pm 1.3 ml/kg/min) exhibited a marginally larger increase in relative VO\textsubscript{2peak} over time compared to the decrease in wave 3 (-3.29 \pm 2.3 ml/kg/min), \(p = .056\); however, paired samples t-tests once again indicated that waves 1 and 3 did not demonstrate a significant change from pre- to post-test, \(t\)'s \leq 1.70, \(p\)'s \geq .11.
Table 2.

*Mean (SE) Values for Participant Demographics Included in Sentence Comprehension and ERP Analyses.*

<table>
<thead>
<tr>
<th>Measure (SE)</th>
<th>Control</th>
<th>Intervetion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
</tr>
<tr>
<td>n</td>
<td>55 (34 female)</td>
<td>-</td>
</tr>
<tr>
<td>n per Wave (1-4)</td>
<td>16/9/12/18</td>
<td>-</td>
</tr>
<tr>
<td>Race (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>7 (12.7)</td>
<td>-</td>
</tr>
<tr>
<td>Black/African American</td>
<td>8 (14.5)</td>
<td>-</td>
</tr>
<tr>
<td>White/Caucasian</td>
<td>25 (45.5)</td>
<td>-</td>
</tr>
<tr>
<td>Mixed Race or Other</td>
<td>15 (27.3)</td>
<td>-</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4 (7.3)</td>
<td>-</td>
</tr>
<tr>
<td>Age (years)</td>
<td>8.7 (0.1) a</td>
<td>9.4 (0.1) b</td>
</tr>
<tr>
<td>Pubertal Timing</td>
<td>1.3 (0.1) a</td>
<td>1.5 (0.1) b</td>
</tr>
<tr>
<td>ADHD</td>
<td>39.6 (3.5) a</td>
<td>40.9 (3.7) a</td>
</tr>
<tr>
<td>Socioeconomic Status</td>
<td>1.8 (0.1) a</td>
<td>1.7 (0.1) a</td>
</tr>
<tr>
<td>WJ-BIA (IQ)</td>
<td>112.6 (1.6) a</td>
<td>115.1 (1.8) b</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>135.7 (0.9) a</td>
<td>139.1 (0.9) b</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>34.8 (1.3) a</td>
<td>38.3 (1.4) b</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>18.7 (0.6) a</td>
<td>19.6 (0.6) b</td>
</tr>
<tr>
<td>BMI Percentile</td>
<td>66.3 (3.8) a</td>
<td>70.3 (3.5) b</td>
</tr>
<tr>
<td>Relative VO2peak (ml/kg/min)</td>
<td>42.5 (1.0) a</td>
<td>42.1 (1.0) a</td>
</tr>
<tr>
<td>VO2max Percentile</td>
<td>39.8 (4.2) a</td>
<td>37.7 (4.1) a</td>
</tr>
</tbody>
</table>

*Note.* Values sharing a common superscript are not statistically different at α = 0.05. ADHD – attention-deficit / hyperactivity disorder; WJ-BIA – Woodcock Johnson Brief Intellectual Ability; IQ – intelligence quotient; BMI – body mass index.
Behavior. In addition to pretest IQ values, SES was entered as a covariate in the following analyses due to differences between control and intervention groups. Mean values for children’s sentence task performance across control and intervention groups are provided in Table 3 (collapsed across wave). Values for any significant differences across waves are reported as (mean ± standard error).

RT. Analyses revealed interactions of Time × Treatment, F(1,109) = 5.74, p = .018, \( \eta_p^2 = .050 \), and Time × Trial × Wave, F(5.8,211.1) = 2.30, p = .037, \( \eta_p^2 = .060 \), which were superseded by a Time × Trial × Treatment × Wave interaction, F(5.8,211.1) = 3.59, p = .002, \( \eta_p^2 = .090 \).

First, Trial × Treatment × Wave RM-ANOVAs were conducted at both pre- and post-test, which revealed main effects of Wave, F’s(3,109) ≥ 3.25, p’s ≤ .025, \( \eta_p^2 \)’s ≥ .082, at each time point. At pre-test, wave 1 (1111.46 ± 37.7 ms) exhibited overall shorter RT, p’s ≤ .04, d’s ≥ 0.79, compared to the other three waves (wave 2: 1558.20 ± 96.1; wave 3: 1515.12 ± 82.5; wave 4: 1364.87 ± 67.8 ms), whereas the only significant difference, p < .04, d = 0.81, that remained at post-test was between waves 1 and 2 (wave 1: 965.60 ± 52.3; wave 2: 1229.10 ± 80.3; wave 3: 1149.36 ± 62.1; wave 4: 1135.01 ± 62.3 ms). No pre- or post-test differences were observed between waves 2, 3, and 4, p’s ≥ .39.

Next, Time × Trial × Treatment RM-MANOVAs were conducted for each wave, which revealed no significant effects for waves 1, 3, or 4, F’s ≤ 2.39, p’s ≥ .13, \( \eta_p^2 \)’s ≤ .072; however, there was a significant three-way interaction for wave 2, F(1.6,29.1) = 6.23, p = .009, \( \eta_p^2 = .257 \).

To break down this interaction, Time × Trial analyses were completed for both the control and intervention groups, which indicated no significant effects, F’s ≤ 3.54, p’s ≥ .08, \( \eta_p^2 \)’s ≤ .371 (a Time × Trial interaction was marginal in the control group). Time × Treatment analyses for each
trial type in wave 2 revealed no effects for congruent and semantic trials, F’s ≤ 1.29, p’s ≥ .27, $\eta_p^2$’s ≤ .067, yet a significant interaction was observed for syntactic trials, F(1,18) = 8.66, p = .009, $\eta_p^2 = .325$. Independent sample t-tests revealed no significant pre- or post-test differences between control (pre: 1245.98 ± 96.3; post: 1298.56 ± 105.4 ms) and intervention (pre: 1392.13 ± 150.0; post: 1002.71 ± 94.9 ms) groups, t’s(20) ≤ 2.06, p’s ≥ .053, but results indicated that the change (control: 52.58 ± 95.7; intervention: -389.42 ± 108.3 ms) in RT was significantly different between the two groups, t(20) = 2.89, p = .009, d = 1.31, such that the intervention group had greater reductions in RT. Paired t-tests further revealed that the control group’s syntactic RT did not differ between pre- and post-test, t(8) = 0.55, p = .60, whereas the intervention group displayed significantly shorter RT over time, t(12) = 3.60, p = .004, d = 0.78. The final analyses for the wave 2 three-way interaction included Trial × Treatment at both pre- and post-test. While no effects were witnessed at post-test, F’s ≤ 1.62, p’s ≥ .22, $\eta_p^2$’s ≤ .082, there was a significant Trial × Treatment interaction at pre-test, F(1.9,34.5) = 5.33, p = .01, $\eta_p^2 = .228$. Independent t-tests indicated that there were no differences between control and intervention groups for any trial type, t’s(20) ≤ 1.11, p’s ≥ .28. Paired samples t-tests revealed that RT for syntactic trials (1245.98 ± 96.3 ms) was significantly shorter compared to congruent (1813.04 ± 156.1 ms) and semantic trials (1624.40 ± 145.7 ms) in the control group, t’s(8) ≥ 3.62, p’s ≤ .007, d’s ≥ 0.94, while congruent and semantic trials were only marginally different, t(8) = 2.10, p = .07. No significant trial differences were witnessed for the intervention group (congruent: 1577.30 ± 139.9; semantic: 1519.34 ± 156.2; syntactic: 1392.13 ± 150.0 ms), t’s(12) ≤ 2.08, p’s ≥ .06.

Returning to the breakdown of the four-way interaction, Time × Treatment × Wave RM-MANOVAs for each trial type revealed a main effect of Wave for all three trials, F’s ≥ 4.78, p’s
\[ \leq .004, \eta^2_p \geq .116, \text{confirming earlier results that wave 1 had overall shorter RT than the other waves. A significant Time \times Treatment interaction was witnessed for syntactic trials, } F(1,109) = 11.10, p = .001, \eta^2_p = .092, \text{ which was marginal for semantic trials as well, } F(1,109) = 3.60, p = .061, \eta^2_p = .032. \text{ Independent t-tests revealed no pre- or post-test differences between groups for either trial type, } t's(117) \leq 1.55, p's \geq .13; \text{ however, the intervention group } (-323.79 \pm 52.2 \text{ ms}) \text{ exhibited larger reductions in RT for syntactic trials compared to the control group } (-127.56 \pm 39.5 \text{ ms}), t(117) = 2.92, p = .004, d = 0.54 (\text{marginal for semantic trials, } p = .064). \text{ Paired samples t-tests confirmed both groups has significantly shorter RT at post-test, } t's \geq 2.80, p's \leq .007, d's \geq 0.38. \]

Lastly, Time \times Trial \times Wave RM-MANOVAs in both the control and intervention groups were not pursued, as comparisons between the two groups were the primary focus of the current study (and significant differences reflected the pattern of results already described above; i.e., wave 1 had shorter RT).

\textit{Accuracy.} A main effect of Time, \( F(1,109) = 4.57, p = .035, \eta^2_p = .040, \) was superseded by a Time \times Trial \times Wave interaction, \( F(5.4,197.7) = 4.04, p = .001, \eta^2_p = .100. \) First, Time \times Trial RM-MANOVAs conducted for each wave revealed no significant effects, \( F's \leq 2.35, p's \geq .14, \eta^2_p 's \leq .083. \) Next, Trial \times Wave MANOVAs at each time point demonstrated no significant effects at pre-test, \( F's \leq 1.25, p's \geq .30, \eta^2_p 's \leq .032, \) yet a significant Trial \times Wave interaction was observed at post-test, \( F(5.6,212.9) = 3.67, p = .002, \eta^2_p = .089. \) ANOVAs comparing waves for each trial type revealed a significant effect for congruent trials, \( F(3,118) = 2.95, p = .036, \) suggesting wave 3 (87.72 ± 1.4) had higher post-test accuracy compared to wave 2 (80.41 ± 2.5); however, Welch’s statistic was not significant after accounting for unequal variances, \( p = .11. \)
Paired samples t-tests comparing trials within each wave further indicated that post-test congruent (80.41 ± 1.4) accuracy was significantly lower compared to semantic (90.18 ± 1.3) and syntactic trials (90.09 ± 1.4) in wave 2, $t$'s(21) ≥ 3.22, $p$'s ≤ .004, $d$'s ≥ 1.03, whereas the same relationship between congruent (86.51 ± 1.7) and semantic (89.86 ± 1.4) trials was marginal in wave 4, $t$(34) = 1.44, $p$ = .058. No significant trial differences were observed in waves 1 or 3, $t$'s ≤ 1.31, $p$’s ≥ .20. Lastly, Time × Wave RM-MANOVAs for each trial revealed a main effect of Time, $F$(1,113) = 4.28, $p$ = .041, $\eta^2_p$ = .036, for syntactic trials suggesting all children increased performance over time. A significant Time × Wave interaction, $F$(3,113) = 3.95, $p$ = .01, $\eta^2_p$ = .095, for congruent trials was also witnessed. One-way ANOVAs comparing waves once again revealed the non-significant post-test difference between waves 2 and 3 reported earlier, as well as a significant effect for change in congruent accuracy, $F$(3,115) = 4.34, $p$ = .026, suggesting waves 1 (4.63 ± 2.4), 3 (3.49 ± 1.5), and 4 (2.95 ± 1.8) exhibited greater increases in accuracy compared to wave 2 (-5.97 ± 2.8), $d$’s ≥ 0.74. Paired samples t-tests further revealed that changes in congruent accuracy from pre- to post-test were marginal for waves 1-3 following Bonferroni correction, $t$’s ≤ 2.39, $p$’s ≥ .024, and non-significant for wave 4, $t$ = 1.65, $p$ = .11.
Table 3.

Mean (SE) Values for Sentence Task Behavior.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Control Pre-test</th>
<th>Post-test</th>
<th>Change</th>
<th>Intervention Pre-test</th>
<th>Post-test</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction Time (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruent</td>
<td>1415.4 (59.7)</td>
<td>1204.4 (50.0)</td>
<td>-211.0 (47.6)</td>
<td>1445.0 (59.7)</td>
<td>1115.0 (46.3)</td>
<td>-330.0 (53.0)</td>
</tr>
<tr>
<td>Semantic</td>
<td>1356.8 (61.1)</td>
<td>1190.0 (58.9)</td>
<td>-166.8 (59.7)</td>
<td>1425.6 (56.6)</td>
<td>1097.1 (48.3)</td>
<td>-328.5 (61.5)</td>
</tr>
<tr>
<td>Syntactic</td>
<td>1164.5 (49.8)</td>
<td>1037.0 (49.3)</td>
<td>-127.6 (39.5)</td>
<td>1264.2 (54.4)</td>
<td>940.4 (39.5)</td>
<td>-323.8 (52.2)</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruent</td>
<td>84.5 (1.2)</td>
<td>85.0 (1.4)</td>
<td>0.5 (1.5)</td>
<td>82.8 (1.4)</td>
<td>85.9 (1.1)</td>
<td>3.1 (1.6)</td>
</tr>
<tr>
<td>Semantic</td>
<td>84.2 (1.5)</td>
<td>87.9 (1.2)</td>
<td>3.6 (1.3)</td>
<td>84.9 (1.2)</td>
<td>89.7 (1.1)</td>
<td>4.8 (1.5)</td>
</tr>
<tr>
<td>Syntactic</td>
<td>84.5 (1.1)</td>
<td>88.1 (1.1)</td>
<td>3.5 (1.3)</td>
<td>82.1 (1.4)</td>
<td>86.9 (1.1)</td>
<td>4.8 (1.6)</td>
</tr>
</tbody>
</table>

Note. Values sharing a common symbol/superscript are not statistically different at $\alpha = 0.05$. 
**ERPs.** A RM-MANOVA for number of accepted ERP trials revealed an expected main effect of Trial, $F(1.4,161.2) = 2321.19, p < .001, \eta^2_p = .954$, which indicated that there was a greater number of congruent trials (M = 55.93, SE = 0.7) compared to semantic (M = 28.8, SE = 0.4) and syntactic (M = 28.0, SE = 0.4) sentences; however, there were no significant treatment group differences for the number of accepted ERP trials, F’s ≤ 2.32, p’s ≥ .08, $\eta^2_p$’s ≤ .059. Mean ERP component values across control and intervention groups are provided in Table 4 (collapsed across wave). Values for any significant differences across waves are reported as (mean ± standard error).

*N400 effect.* There were no significant findings for the N400 effect except a marginal effect of Time, $F(1,109) = 3.40, p = .068, \eta^2_p = .030$, suggesting the magnitude of N400 effect increased over time. The absence of a Trial effect indicated that the magnitude of the amplitude difference between congruent sentences and each of the violation trial types were similar (within the 9-electrode ROI; Figure 1), and that this pattern was consistent across time and treatment groups (Figure 2).

**Figure 1.**

*Note.* Topographic plots of the N400 effect across all participants between 300-500ms (semantic/syntactic - congruent mean amplitude).
**N400 latency.** No significant effects were observed for N400 latency, $F’s \leq 2.46$, $p’s \geq .09, \eta^2_p’s \leq .062$.

**N400 amplitude.** A main effect of Wave, $F(3/109) = 3.22$, $p = .026, \eta^2_p = .081$, indicated that children in wave 1 ($-7.34 \pm 0.7 \mu V$) had overall larger N400 amplitude compared to children in wave 3 ($-4.62 \pm 0.8 \mu V$) and 4 ($-4.62 \pm 0.7 \mu V$), $p’s \leq .011$, $d’s \geq 0.67$, but only marginally larger compared to wave 2 ($-5.42 \pm 0.9 \mu V$), $p = .094$. Children’s N400 amplitude in waves 2-4 did not differ, $p’s \geq .48$. A main effect of Trial, $F(2.0/216.4) = 55.92$, $p < .001, \eta^2_p = .323$, indicated that lexical items in congruent sentences had smaller amplitude (i.e., less negative) compared to those in semantic and syntactic sentences, $t’s(118) \geq 8.05$, $p’s \leq .001, d’s \geq 0.76$, whereas syntactic amplitude was smaller but not significantly different from semantic trials following Bonferroni correction, $t(118) = 2.27$, $p = .025$, $d = 0.21$. No treatment group differences were observed at either time point, $t’s(117) \leq 1.76$, $p’s \geq .08$.

**P600 effect.** No significant effects were observed for the P600 effect, $F’s \leq 2.02$, $p’s \geq .12, \eta^2_p’s \leq .053$. 

88
Table 4.

*Mean (SE) Values for Sentence Task ERP Measures.*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Control Pre-test</th>
<th>Control Post-test</th>
<th>Control Change</th>
<th>Intervention Pre-test</th>
<th>Intervention Post-test</th>
<th>Intervention Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>N400 Latency (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Congruent</em> †</td>
<td>396.42 (2.6) a</td>
<td>394.11 (2.6) a</td>
<td>-2.31 (2.3) a</td>
<td>394.79 (2.2) a</td>
<td>395.43 (2.1) a</td>
<td>0.65 (2.9) a</td>
</tr>
<tr>
<td><em>Semantic</em> †</td>
<td>399.84 (2.4) a</td>
<td>398.98 (2.5) a</td>
<td>-0.86 (2.8) a</td>
<td>401.82 (2.1) a</td>
<td>399.76 (2.2) a</td>
<td>-2.06 (3.3) a</td>
</tr>
<tr>
<td><em>Syntactic</em> †</td>
<td>400.48 (2.7) a</td>
<td>394.90 (2.6) a</td>
<td>-5.58 (3.4) a</td>
<td>398.89 (2.3) a</td>
<td>398.75 (2.4) a</td>
<td>-0.14 (3.2) a</td>
</tr>
<tr>
<td>N400 Amplitude (μV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Congruent</em> ‡</td>
<td>-3.24 (0.6) a</td>
<td>-3.11 (0.7) a</td>
<td>0.13 (0.6) a</td>
<td>-4.80 (0.6) a</td>
<td>-3.51 (0.5) a</td>
<td>1.29 (0.7) a</td>
</tr>
<tr>
<td><em>Semantic</em> ‡</td>
<td>-6.26 (0.8) a</td>
<td>-6.75 (0.8) a</td>
<td>-0.50 (0.8) a</td>
<td>-7.38 (0.7) a</td>
<td>-6.83 (0.7) a</td>
<td>0.54 (0.8) a</td>
</tr>
<tr>
<td><em>Syntactic</em> ‡</td>
<td>-5.37 (0.8) a</td>
<td>-5.70 (0.8) a</td>
<td>-0.33 (0.7) a</td>
<td>-6.70 (0.7) a</td>
<td>-6.48 (0.7) a</td>
<td>0.22 (0.7) a</td>
</tr>
<tr>
<td>N400 Effect (μV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sem. - Cong. Diff †</td>
<td>-3.14 (0.6) a</td>
<td>-3.90 (0.7) a</td>
<td>-0.75 (1.0) a</td>
<td>-3.08 (0.6) a</td>
<td>-2.92 (0.6) a</td>
<td>0.15 (0.8) a</td>
</tr>
<tr>
<td>Syn. - Cong. Diff †</td>
<td>-2.65 (0.5) a</td>
<td>-3.00 (0.7) a</td>
<td>-0.35 (0.8) a</td>
<td>-2.74 (0.6) a</td>
<td>-3.42 (0.5) a</td>
<td>-0.68 (0.7) a</td>
</tr>
<tr>
<td>P600 Effect (μV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sem. - Cong. Diff †</td>
<td>-0.34 (0.7) a</td>
<td>-0.90 (0.8) a</td>
<td>-0.56 (1.0) a</td>
<td>-1.29 (0.7) a</td>
<td>0.40 (0.7) a</td>
<td>1.67 (0.9) a</td>
</tr>
<tr>
<td>Syn. - Cong. Diff †</td>
<td>0.94 (0.8) a</td>
<td>1.17 (0.7) a</td>
<td>0.23 (1.1) a</td>
<td>0.63 (0.7) a</td>
<td>0.67 (0.6) a</td>
<td>0.03 (0.8) a</td>
</tr>
</tbody>
</table>

*Note.* Measures/values sharing a common symbol/superscript are not statistically different at α = 0.05.
Figure 2.

![Amplitude vs Time Graphs](image)

**Note.** Pre- and post-test grand average waveforms from the 9-electrode (Cz, CPz, Pz, C1/2, CP1/2, and P1/2) region of interest. Congruent trials are compared to semantic violations on the left, and syntactic violations on the right. No group differences were observed between children in the control (row A) and intervention groups (row B).

**Changes in Aerobic Fitness and Academic Achievement**

Given that the intervention and control groups displayed varying levels of fitness change across each wave, children were collapsed across treatment/wave and hierarchical regression analyses were performed to determine if changes in aerobic fitness were related to children’s post-test academic performance after controlling for pre-test scores, demographic/fitness variables (age, sex, SES, IQ, BMI percentile, VO2max percentile), as well as treatment/wave assignment. Further, the relationship between ERPs and academic performance was investigated.
to determine: 1) if specific ERP components/measures were independently related to academics after controlling for important demographics, and 2) if significant ERP measures potentially mediated any significant relationships between changes in aerobic fitness and academic performance. Given the earlier ERP findings that revealed no significant differences between violation trials for their overall N400 amplitude/latency, N400 effect size, or P600 effect size, these measures were collapsed across semantic and syntactic trials for all subsequent analyses and referred to as violation trials. Also, considering there were no significant ERP differences observed across time, ERP measures and academics were investigated using pre-test values.

**ERPs to academics.**

**Bivariate correlations.**

*Demographics & academics.* Pre-test Pearson correlations (all p’s < .05) revealed no significant relations between academic measures and age, SES, ADHD, or aerobic fitness. IQ was positively related to each academic subtest and composite score, r’s ≤ .34, indicating children with higher IQ exhibited better academic achievement. Sex was negatively related to both reading comprehension and written expression subtests, r’s ≤ -.22, suggesting females outperformed males. As for BMI percentile, negative associations were observed for math subtests and composite scores, word recognition and reading composite, as well as overall achievement, r’s ≤ -.19; thus, children with higher BMI percentiles demonstrated lower academic scores.

*Demographics & ERPs.* Correlations indicated that age, SES, and IQ were related to N400 amplitude for violation trials, r’s ≤ -.19, suggesting that older children, those with higher IQ, and children residing in higher SES households had larger (i.e., more negative) amplitude. VO$_{2\text{max}}$ percentile was positively related to congruent N400 latency, r = .19, whereas IQ was
negatively related, \( r = -0.23 \), such that higher fit children and those with lower IQ had longer latencies.

**Academics & ERPs.** Table 5 provides Pearson correlation coefficients between children’s standardized academic scores and the six primary ERP measures: overall N400 amplitude/latency for congruent and violation trials, as well as the N400/P600 effect amplitude for violation trials. As shown, N400 amplitude for violation trials and the P600 effect exhibited several significant associations with academic measures that primarily involved aspects of language processing, suggesting that larger overall N400 amplitude (as indexed by larger negative values) and a larger (positive) P600 effect were related to better academic achievement. Neither ERP measure was associated with math computation or composite scores, nor were they related to reading comprehension (although N400 amplitude approached significance, \( p = 0.06 \)). Compared to the P600 effect, larger N400 amplitude was selectively related to math concepts and listening comprehension, whereas the P600 effect was related to written expression. The only other significant associations were observed for congruent N400 latency and the N400 effect with listening comprehension, suggesting that shorter latency and greater amplitude reductions for congruent trials (i.e., larger negative effect values) were related to better comprehension.

**Hierarchical regression.** Table 6 provides a summary of the analyses regressing pre-test academic performance on significant demographic variables and intervention classification in Step 1, wave in Step 2, and ERP measures in Step 3. Age and SES were entered in Step 1 due to their association with N400 amplitude for violation trials, whereas IQ and BMI percentile were entered due to significant correlations with academic achievement. Sex was included in Step 1 for reading comprehension and written expression subtests only. As for the ERP measures
included in Step 3, N400 amplitude for violation trials and the P600 effect were entered into the model together to determine if both measures were independently related to a particular academic domain, or if a single measure accounted for a majority of the variance. Congruent N400 latency and the N400 effect were also entered in Step 3, but for listening comprehension only.

Step 1 indicated that higher IQ levels were associated with better performance for each academic measure (partial correlation: \[pr\]'s ≥ .35, \(p\)'s ≤ .001), while older children exhibited greater performance for word recognition fluency, fluency composite scores, written expression and composite scores, as well as math concepts (\(pr\)'s ≥ .17, \(p\)'s ≤ .05). Sex was related to reading comprehension, and written expression scores (\(pr\)'s ≤ -.17, \(p\)'s ≤ .04), suggesting females outperformed males. BMI percentile was independently associated with reading composite scores and subtests, written composite, math concepts/composite, and overall achievement (\(pr\)'s ≤ -.19, \(p\)'s ≤ .05), indicating that children with higher BMI percentiles had lower performance.

No significant differences were observed between the intervention and control groups, \(t\)'s ≤ 1.27, \(p\)'s ≥ .21, except for a marginal effect for listening comprehension, \(t = 1.94, p = .055, pr = .18\), such that children in the intervention group had higher pre-test scores. Step 2 including wave was not significant for any of the academic measures, \(\Delta R^2\)'s ≤ .043, \(p\)'s ≥ .09. As for the ERP measures, Step 3 revealed that greater N400 amplitude (\(pr\)'s ≤ -.18, \(p\)'s ≤ .05) and a larger P600 effect (\(pr\)'s ≥ .26, \(p\)'s ≤ .01) for violations trials were both related to better performance on fluency subtests and composite scores, word recognition, reading composite, and spelling. Additionally, a larger P600 effect was selectively associated with greater written composite scores, listening comprehension, and overall achievement (\(pr\)'s ≥ .24, \(p\)'s ≤ .015). Lastly, a larger
N400 effect was related to better listening comprehension ($pr = -.23, p < .02$), whereas congruent latency was not significant ($pr = -.09, p = .36$).
Table 5.

*Bivariate Correlations between ERP Measures and Academic Performance.

<table>
<thead>
<tr>
<th>Academic Measure</th>
<th>N400 Amplitude</th>
<th></th>
<th></th>
<th>N400 Latency</th>
<th></th>
<th></th>
<th>N400 Effect</th>
<th>P600 Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Congruent</td>
<td>Sem./Syn.</td>
<td>Congruent</td>
<td>Sem./Syn.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Composite</td>
<td>-.05</td>
<td>-.18*</td>
<td>-.06</td>
<td>-.08</td>
<td>-.09</td>
<td>.23*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>-.01</td>
<td>-.17</td>
<td>-.04</td>
<td>-.05</td>
<td>-.13</td>
<td>.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Recognition</td>
<td>-.11</td>
<td>-.20*</td>
<td>-.05</td>
<td>-.06</td>
<td>-.05</td>
<td>.21*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency Composite</td>
<td>-.17</td>
<td>-.29**</td>
<td>.02</td>
<td>-.02</td>
<td>.02</td>
<td>.25*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Recognition</td>
<td>-.16</td>
<td>-.30**</td>
<td>.04</td>
<td>-.09</td>
<td>-.02</td>
<td>.25*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonword Decoding</td>
<td>-.18</td>
<td>-.25*</td>
<td>.08</td>
<td>.05</td>
<td>.06</td>
<td>.22*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Written Composite</td>
<td>-.09</td>
<td>-.21*</td>
<td>.06</td>
<td>-.08</td>
<td>.03</td>
<td>.28*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spelling</td>
<td>-.10</td>
<td>-.21*</td>
<td>.06</td>
<td>-.09</td>
<td>.01</td>
<td>.27*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Written Expression</td>
<td>-.03</td>
<td>-.12</td>
<td>.03</td>
<td>-.04</td>
<td>.04</td>
<td>.19*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math Composite</td>
<td>-.01</td>
<td>-.11</td>
<td>-.10</td>
<td>-.04</td>
<td>-.03</td>
<td>.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concepts</td>
<td>-.10</td>
<td>-.21*</td>
<td>-.11</td>
<td>-.05</td>
<td>-.02</td>
<td>.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computation</td>
<td>.07</td>
<td>.00</td>
<td>-.07</td>
<td>-.02</td>
<td>-.02</td>
<td>.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening Comprehension</td>
<td>-.09</td>
<td>-.29**</td>
<td>-.20*</td>
<td>-.11</td>
<td>-.18*</td>
<td>.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Achievement</td>
<td>-.06</td>
<td>-.22*</td>
<td>-.11</td>
<td>-.07</td>
<td>-.08</td>
<td>.20*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p ≤ .05, ** p ≤ .001.
Table 6.

Pre-test Academic Performance Regression Values for Violation Trial N400 Amplitude and P600 Effect.

<table>
<thead>
<tr>
<th>Academic Measure</th>
<th>Step 1 R²</th>
<th>Step 3 AR²</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading Composite</td>
<td>.335**</td>
<td>.060*</td>
<td>-.448/.726</td>
<td>.231/.224</td>
<td>-.172/.261</td>
<td>1.94*/3.24*</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>.318**</td>
<td>.021</td>
<td>-.262/.445</td>
<td>.247/.240</td>
<td>-.099/.157</td>
<td>1.06/1.86</td>
</tr>
<tr>
<td>Word Recognition</td>
<td>.250**</td>
<td>.074*</td>
<td>-.547/.663</td>
<td>.216/.210</td>
<td>-.239/.270</td>
<td>2.53*/3.15*</td>
</tr>
<tr>
<td>Fluency Composite</td>
<td>.214**</td>
<td>.117**</td>
<td>-.873/.932</td>
<td>.249/.242</td>
<td>-.326/.325</td>
<td>3.50**/3.85**</td>
</tr>
<tr>
<td>Word Recognition</td>
<td>.244**</td>
<td>.104**</td>
<td>-.823/.930</td>
<td>.251/.244</td>
<td>-.298/.314</td>
<td>3.28**/3.81**</td>
</tr>
<tr>
<td>Nonword Decoding</td>
<td>.148*</td>
<td>.104**</td>
<td>-.814/.833</td>
<td>.257/.250</td>
<td>-.316/.301</td>
<td>3.17*/3.34**</td>
</tr>
<tr>
<td>Written Composite</td>
<td>.275**</td>
<td>.076*</td>
<td>-.464/.881</td>
<td>.256/.249</td>
<td>-.169/.299</td>
<td>1.81/3.54**</td>
</tr>
<tr>
<td>Spelling</td>
<td>.168**</td>
<td>.118**</td>
<td>-.837/1.148</td>
<td>.290/.281</td>
<td>-.277/.355</td>
<td>2.89*/4.08**</td>
</tr>
<tr>
<td>Written Expression</td>
<td>.305**</td>
<td>.022</td>
<td>.080/.406</td>
<td>.263/.255</td>
<td>.029/.138</td>
<td>0.30/1.59</td>
</tr>
<tr>
<td>Math Composite</td>
<td>.333**</td>
<td>.005</td>
<td>.102/.204</td>
<td>.314/.305</td>
<td>.030/.057</td>
<td>0.32/0.67</td>
</tr>
<tr>
<td>Concepts</td>
<td>.388**</td>
<td>.015</td>
<td>-.196/.428</td>
<td>.263/.255</td>
<td>-.067/.136</td>
<td>0.74/1.67</td>
</tr>
<tr>
<td>Computation</td>
<td>.231**</td>
<td>.009</td>
<td>.381/.066</td>
<td>.348/.338</td>
<td>.110/.018</td>
<td>1.10/0.20</td>
</tr>
<tr>
<td>Listening Comprehension</td>
<td>.281**</td>
<td>.080*</td>
<td>-.357/.784</td>
<td>.233/.278</td>
<td>-.147/.301</td>
<td>1.53/2.82*</td>
</tr>
<tr>
<td>Overall Achievement</td>
<td>.494**</td>
<td>.028*</td>
<td>-.264/.529</td>
<td>.216/.209</td>
<td>-.098/.183</td>
<td>1.23/2.53*</td>
</tr>
</tbody>
</table>

Note. Treatment group and demographic variables (i.e., age, sex, socioeconomic status, BMI [body mass index] percentile, IQ [intelligence quotient]) were included in Step 1 (if significantly correlated with the dependent academic outcome), while N400 amplitude for violation trials (left-side column values) and the P600 effect (right-side column values) were entered in Step 3. Step 2 included wave, which was not significant for any of the academic measures, ΔR²’s ≤ .043, p’s ≥ .09.

* p ≤ .05, ** p ≤ .001.
Changes in aerobic fitness.

Bivariate correlations. Pearson correlations (all $p$’s < .05) indicated that changes in VO$_{2\text{max}}$ percentile were positively related to children’s post-test academic performance for both reading subtests and reading composite scores, as well as overall composite achievement scores, $r$’s $\geq .19$. Changes in aerobic fitness were not related to changes in any ERP measure, $p$’s $\geq .38$, nor were changes in ERPs related to changes in academic performance, $p$’s $\geq .09$; thus, mediation analyses incorporating changes in fitness and ERPs were not necessary. Rather, significant pre-test ERP measures identified in the previous set of analyses were included in Step 1 of the analyses regressing post-test academic performance on changes in fitness (Step 3). Step 1 also controlled for intervention assignment, pretest academic performance and fitness levels, in addition to significant pre-test demographics identified in prior analyses, while wave was entered into Step 2.

Hierarchical regression. Despite significant correlations between increases in aerobic fitness and greater academic performance at post-test, particularly for reading scores, these relationships were only marginally significant ($pr$’s $\geq .15$, $p$’s $\leq .09$) after controlling for pretest demographics, academic performance, and ERPs, as well as a number of significant differences between waves (as highlighted by the earlier RM-MANOVA findings). No other marginal relationships were observed between increases in fitness and improvements in other academic areas.
Lower versus Higher Fit Individuals

The final analyses were conducted to replicate prior work (Scudder et al., 2014a) comparing children residing at the lower (≤ 30th percentile) and higher (≥ 70th percentile) ends of the fitness distribution; however, over half of the children included in the current ERP analyses were classified as lower (n = 63) versus higher fit (n = 28). As such, pre-test academic performance and sentence behavior/ERPs were compared between evenly divided groups of higher and lower fit children that were matched for age, sex, SES, and IQ.

Participants and academic outcomes. Mean values for children’s demographics are provided in Table 7. Independent samples t-tests indicated that children in both groups did not differ on age, pubertal timing, ADHD, SES, or IQ, t’s(54) ≤ 1.06, p’s ≥ .30; however, lower fit children were taller, weighed more, and had larger BMI values/percentiles and lower aerobic fitness levels, t’s(54) ≥ 2.48, p’s ≤ .02, d’s ≥ 0.68. Mean values for children’s standardized academic performance scores are provided in Table 8. The largest group differences were observed for written academic measures, such that higher fit children performed better on written expression, whereas lower fit children tended to have higher spelling scores; however, independent samples t-tests indicated that there were no significant differences between groups for any of the academic outcomes, t’s(54) ≤ 1.41, p’s ≥ .16.
**Table 7.**

*Mean (SE) Values for Demographics between Lower and Higher Fit Participants.*

<table>
<thead>
<tr>
<th>Measure (SE)</th>
<th>Lower Fit</th>
<th>Higher Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>28 (16 female)</td>
<td>28 (16 female)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>8.7 (0.1)</td>
<td>8.7 (0.1)</td>
</tr>
<tr>
<td>ADHD</td>
<td>42.7 (5.8)</td>
<td>43.4 (5.9)</td>
</tr>
<tr>
<td>Socioeconomic Status</td>
<td>2.0 (0.2)</td>
<td>2.2 (0.1)</td>
</tr>
<tr>
<td>WJ-BIA (IQ)</td>
<td>111.2 (2.1)</td>
<td>112.9 (2.0)</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>138.8 (1.4)</td>
<td>134.4 (1.1)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>41.8 (2.0)</td>
<td>29.6 (0.8)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>21.5 (0.8)</td>
<td>16.3 (0.3)</td>
</tr>
<tr>
<td>BMI Percentile</td>
<td>85.1 (3.8)</td>
<td>51.8 (4.5)</td>
</tr>
<tr>
<td>Relative VO₂peak (ml/kg/min)</td>
<td>36.0 (1.0)</td>
<td>52.0 (0.8)</td>
</tr>
<tr>
<td>VO₂max Percentile</td>
<td>10.0 (1.1)</td>
<td>84.5 (0.9)</td>
</tr>
</tbody>
</table>

*Note.* Values sharing a common superscript are not statistically different at $\alpha = 0.05$. ADHD – attention-deficit / hyperactivity disorder; WJ-BIA – Woodcock Johnson Brief Intellectual Ability; IQ – intelligence quotient; BMI – body mass index.

**Sentence task performance and ERPs.**

*Behavior.* Mean values for children’s sentence task performance are provided in Table 9.

For RT, a main effect of Trial, $F(1.9,104.6) = 28.23, p < .001, \eta^2_p = .343$, indicated that syntactic trials resulted in shorter RT compared to congruent and semantic sentences, $t’$’s(55) ≥ 5.68, $p$’s < .001, $d$’s ≥ 0.77. Congruent trials resulted in the longest RT, but were not statistically different from semantic trials, $t(55) = 1.92, p = .06$. Although it appeared higher fit children had shorter syntactic RT compared to lower fit children, the Group × Trial interaction only approached significance, $F(1.9,104.6) = 2.71, p = .07, \eta^2_p = .048$. As for accuracy, a main effect of Group, $F(1,54) = 4.46, p < .04, \eta^2_p = .076$, indicated that higher fit children had overall greater accuracy across all three trial types, $d = 0.50$. No other significant effects were observed.
Table 8.

Mean (SE) Academic Standardized Scores between Lower and Higher Fit Participants.

<table>
<thead>
<tr>
<th>Academic Measure</th>
<th>Lower Fit</th>
<th>Higher Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading Composite</td>
<td>115.1 (2.6) a</td>
<td>113.9 (2.5) a</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>110.1 (1.9) a</td>
<td>110.8 (2.7) a</td>
</tr>
<tr>
<td>Word Recognition</td>
<td>115.3 (2.1) a</td>
<td>114.1 (2.4) a</td>
</tr>
<tr>
<td>Fluency Composite</td>
<td>112.6 (2.4) a</td>
<td>111.1 (2.7) a</td>
</tr>
<tr>
<td>Word Recognition</td>
<td>113.3 (2.6) a</td>
<td>112.7 (3.0) a</td>
</tr>
<tr>
<td>Nonword Decoding</td>
<td>110.7 (2.1) a</td>
<td>108.8 (2.4) a</td>
</tr>
<tr>
<td>Written Composite</td>
<td>110.4 (2.4) a</td>
<td>111.4 (2.7) a</td>
</tr>
<tr>
<td>Spelling</td>
<td>112.4 (2.8) a</td>
<td>109.0 (3.0) a</td>
</tr>
<tr>
<td>Written Expression</td>
<td>106.8 (2.5) a</td>
<td>111.7 (2.5) a</td>
</tr>
<tr>
<td>Math Composite</td>
<td>110.3 (3.6) a</td>
<td>112.3 (3.2) a</td>
</tr>
<tr>
<td>Concepts</td>
<td>108.5 (2.8) a</td>
<td>109.1 (2.7) a</td>
</tr>
<tr>
<td>Computation</td>
<td>108.0 (3.8) a</td>
<td>110.6 (3.5) a</td>
</tr>
<tr>
<td>Listening Comprehension</td>
<td>107.5 (2.6) a</td>
<td>110.5 (2.1) a</td>
</tr>
<tr>
<td>Overall Achievement</td>
<td>112.0 (2.6) a</td>
<td>114.4 (2.4) a</td>
</tr>
</tbody>
</table>

Note. Values sharing a common superscript are not statistically different at α = 0.05.

Table 9.

Mean (SE) Values for Sentence Task Performance between Lower and Higher Fit Participants.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Lower Fit</th>
<th>Higher Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction Time (ms)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruent †</td>
<td>1599.1 (101.3) a</td>
<td>1548.9 (96.2) a</td>
</tr>
<tr>
<td>Semantic †</td>
<td>1574.5 (104.5) a</td>
<td>1421.9 (88.0) a</td>
</tr>
<tr>
<td>Syntactic ‡</td>
<td>1400.2 (94.0) a</td>
<td>1161.9 (74.6) a</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruent †</td>
<td>82.0 (2.1) a</td>
<td>84.3 (1.6) b</td>
</tr>
<tr>
<td>Semantic †</td>
<td>80.9 (2.1) a</td>
<td>87.2 (2.2) b</td>
</tr>
<tr>
<td>Syntactic †</td>
<td>80.3 (1.9) a</td>
<td>86.1 (1.6) b</td>
</tr>
</tbody>
</table>

Note. Measures/values sharing a common symbol/superscript are not statistically different at α = 0.05.
ERPs. Mean ERP component values are provided in Table 10, and ERP waveforms are shown in Figure 3. A main effect of Group, $F(1,54) = 6.69$, $p < .02$, $\eta^2_p = .110$, revealed that lower fit children had overall shorter N400 latencies compared to higher fit children, $d = 0.70$. As for N400 amplitude, main effects of Trial, $F(1,54) = 24.07$, $p < .001$, $\eta^2_p = .308$, and Group, $F(1,54) = 5.03$, $p < .03$, $\eta^2_p = .085$, indicated that violation trials resulted in larger N400 amplitude compared to congruent sentences, $d = 0.67$, and that higher fit children had overall larger N400 amplitude compared to lower fit individuals, $d = 0.61$. Lastly, no significant differences were observed between groups for the N400 effect, $t(54) = 1.42$, $p = .16$, whereas the analysis for the P600 effect approached significance, $t(54) = 1.94$, $p = .06$, suggesting lower fit children tended to have larger amplitudes. Groups did not differ in the number of accepted ERP trials (congruent: $55.0 \pm 1.8$; violation: $27.7 \pm 0.9$) for waveform averages, $F(1,54) = 2.32$, $p = .13$.

Table 10.

Mean (SE) Values for Sentence ERPs between Lower and Higher Fit Participants.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Lower Fit</th>
<th>Higher Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>N400 Latency (ms)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruent †</td>
<td>389.22 (3.1)$^a$</td>
<td>402.01 (3.0)$^b$</td>
</tr>
<tr>
<td>Violation †</td>
<td>396.53 (2.8)$^a$</td>
<td>402.04 (3.0)$^b$</td>
</tr>
<tr>
<td>N400 Amplitude ( μV)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruent †</td>
<td>-2.53 (0.7)$^a$</td>
<td>-4.59 (0.9)$^b$</td>
</tr>
<tr>
<td>Violation †</td>
<td>-4.30 (0.6)$^a$</td>
<td>-6.83 (0.8)$^b$</td>
</tr>
<tr>
<td>N400 Effect ( μV)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.85 (0.6)$^a$</td>
<td>-2.97 (0.5)$^a$</td>
<td></td>
</tr>
<tr>
<td>P600 Effect ( μV)</td>
<td>1.68 (0.8)$^a$</td>
<td>-0.68 (0.9)$^a$</td>
</tr>
</tbody>
</table>

Note. Measures/values sharing a common symbol/superscript are not statistically different at $\alpha = 0.05$.
Figure 3.

Note. Pre-test grand average waveforms from the 9-electrode (Cz, CPz, Pz, C1/2, CP1/2, and P1/2) region of interest. Congruent and violation sentences are compared in both lower (LF) and higher fit (HF) children.
CHAPTER 5: DISCUSSION

Contrary to previous findings, children participating in the current Fitness Improves Thinking in Kids (FITKids2) after-school physical activity (PA) trial did not exhibit greater improvements in aerobic fitness levels (Hillman et al., 2014) or academic performance compared to the wait-list control group; however, the intervention did appear to be effective for attenuating increases in body weight and body mass index (BMI) over the course of the school year. Similarly, no group differences were observed for the N400 or P600 event-related brain potential (ERP) components, which were recorded during a sentence comprehension task to provide additional information about children’s semantic (i.e., meaning) processing and access to word-related knowledge, as well as their ability to detect syntactic ambiguities and allocate resources towards re-analysis and repair. Interestingly, the control group in wave 1 was the only set of children that exhibited a significant increase in aerobic fitness, which was intriguing given the number of Time × Wave interactions that demonstrated wave 1 had greater post-test academic performance and/or larger improvements over time compared to other participants. Although these interactions did not contain the additional factor of Treatment, it is worth noting the marginally significant three-way interaction for academic composite scores that revealed greater improvements among the wave 1 control group compared to children in the intervention. Wave 1 also demonstrated overall shorter reaction time (RT) and greater improvements in congruent accuracy for the sentence comprehension task, as well as overall larger N400 amplitude compared to the other waves. Such differences are important to highlight given the secondary hierarchical regression analyses that indicated larger N400 and P600 amplitude for violation trials was independently related to several academic tests involving aspects of language processing. As for the association between changes in aerobic fitness and academics (after
collapsing across group/wave), regression analyses revealed that increases in children’s aerobic fitness were marginally related to greater improvements in reading achievement after controlling for pre-test fitness levels and academic scores, significant demographic variables (i.e., age, sex, socioeconomic status [SES], body mass index [BMI] percentile, and IQ [intelligence quotient]), as well as children’s N400 and P600 amplitude for violation trials. Finally, exploratory analyses partially replicated previous findings (Scudder et al., 2014a) by comparing children residing at the lower (≤ 30th percentile) and higher (≥ 70th percentile) ends of the fitness distribution that were matched on age, sex, SES, and IQ. Despite no group differences for any academic outcome, higher fit children demonstrated greater sentence task accuracy and greater overall N400 amplitude compared to lower fit children, whereas lower fit children displayed shorter N400 latency.

**After-School Physical Activity Program**

Previous research has demonstrated that school-based (Carrel et al., 2005; Dobbins et al., 2013) and after-school (Hillman et al., 2014) PA programs are often effective strategies to increases children’s physical activity and aerobic fitness levels. Therefore, it was surprising to find that children who participated in the current Fitness Improves Thinking in Kids (FITKids) program did not exhibit greater aerobic fitness gains compared to their peers in the wait-list control group. Findings from the initial FITKids study indicated that the intervention group had larger increases in VO2max percentile (~ 5.6%) compared to the wait-list control (0.8%), yet it is important to note that children’s average fitness percentile was ~ 21% whereas fitness levels in the current study were ~ 15% higher (in either case both groups had rather low aerobic fitness levels). As such, one possible reason for the lack of fitness change could have been that children in the current study were higher fit, and may have required activities that incorporated higher
intensities to produce comparable fitness changes; however, what remains unknown are the potential reasons for the small but significant increase in aerobic fitness among the wave 1 control group. One of the limitations of a waitlist control group is that it is difficult to account for children’s PA and behavior outside the program, and unfortunately objective PA measurement was not used in the current study; thus, it is unclear if these children participated in other physically demanding activities during the day, or if children’s PA levels were even related to changes in aerobic fitness. Despite the unexpected fitness differences across wave/group, it was interesting to note that the control group had significantly larger increases in BMI percentile among the entire sample. Although it appeared as if the intervention group had larger BMI values at pre-test, analyses revealed that there were no group differences at either time point, and groups did not differ on measures of pubertal timing. Therefore, as opposed to the notion that the control group simply caught up, the PA program may have been successful in maintaining body composition levels over the course of the academic year, as demonstrated in previous research (Donnelly et al., 2009). Regarding the other demographic variables, no group differences were observed at pre- or post-test with the exception of IQ, such that the control group had higher pre-test values. Unfortunately, this was a major limitation given that IQ was by far the most significant predictor of children’s academic achievement.

Another unexpected finding was that the intervention group did not exhibit greater improvements in academic performance than the control group. Instead, it was revealed that children in wave 1 demonstrated superior performance and larger improvements for a number of academic tests compared to the other waves, with evidence further suggesting that the wave 1 control group (i.e., the only group that increased aerobic fitness) exhibited greater increases for overall academic composite scores than children in the intervention. As for the sentence
comprehension task, children performed at a high level of accuracy on average (> 80%), which suggested that all participants were able to adequately read and identify mistakes. Significant performance differences were once again observed between waves, with wave 1 exhibiting overall shorter RT and larger improvements for congruent accuracy (significantly greater than wave 4). ERP measures that were recorded during sentence reading replicated a well-established pattern of findings throughout the language processing literature, including a previous study comparing higher and lower fit individuals using the same sentence paradigm (Scudder et al., 2014a). That is, children exhibited smaller (i.e., less negative) N400 amplitude for words presented in a facilitative/congruent context, compared to sentences in which a semantic or syntactic violation occurred. As is shown in Figure 2, groups did not differ on the size of this robust amplitude difference between congruent and violation sentences (i.e., the N400 effect), or for N400 amplitude/latency on any of the individual trial types; however, analyses did reveal that children in wave 1 had overall larger N400 amplitude compared to the other waves. It is also shown that both groups demonstrated a negligible P600 effect for syntactic sentences that contained a word-order violation. Such a result was not entirely unexpected, especially considering an earlier finding that only higher fit individuals (≥ 70th percentile) displayed a distinct P600 effect among a sample of children that were slightly older than those in the current study (Scudder et al., 2014a). Previous studies in adults (Aydelott et al., 2006; Neville et al., 1991) have also witnessed robust P600 effects in response to word-order violations; however, it may have been that the younger children included in the current study are still developing the necessary skills/abilities to accurately identify such mistakes as syntactically ambiguous. After all, certain ERP components that have been linked to aspects of syntactic processing, such as the (Early) Left Anterior Negativity (ELAN and LAN), do not reliably occur until the teenage years
(Hahne et al., 2004), suggesting that these processes continue to mature well into adolescence. As evidenced by the presence of a large N400 effect to syntactic sentences, children may have partially relied on semantic processing for the identification of anomalies/mistakes within the context of each sentence. Future studies would benefit from the inclusion of additional types of syntactic mistakes that are perhaps more salient or easier to identify when reading word-by-word, such as verb tense violations (“My uncle will watching the movie”; Silva-Pereyra et al., 2005).

**ERPs to Academics: N400 and P600 Amplitude**

A secondary aim of the current study was to determine whether ERP indices of language processing potentially mediate any observed relationships between fitness and academics. Initial group analyses revealed that children’s amplitude/latency for both the N400 and P600 components remained relatively unchanged across time, and bivariate correlations revealed no significant relation between changes in ERPs and academic performance. However, when pre-test academic scores were regressed onto children’s N400 and P600 amplitude, larger amplitude for both components was independently related to better performance for several academic outcomes specifically involving aspects of language/reading (no relationships were observed for math). Whereas greater amplitude for both components was associated with better performance on word recognition, spelling, listening comprehension, and fluency subtests/composite scores, a larger P600 effect was selectively related to higher scores for reading and written composite, as well as overall achievement. A larger N400 effect was also related to better listening comprehension, yet no significant relationships were observed for N400 latency. These findings replicate previous findings in children that have revealed larger N400 amplitudes among higher ability readers (Coch & Holcomb, 2003) and individuals with greater vocabulary (Khalifian et
al., in press), and extend this relationship to the P600 component and other language-based academic abilities. Given these significant associations with academic achievement, even after controlling for important factors such as IQ, ERP components appear to be valuable indexes of children’s academic performance that may hold future promise as biomarkers for predicting children’s achievement in school. Comparing ERPs over the course of development is made difficult by the growth and maturation of the skull/brain; however, tracking these measures over longer periods of time will be necessary to determine if/how potential component modulations are related to changes in children’s academic performance or aspects of cardiovascular health.

It was also interesting to note that children’s academic performance was related to N400 amplitude for violation trials in particular. It is important to keep in mind that the supportive content/structure of congruent sentences facilitated semantic processing and resulted in decreased N400 amplitude, depending on the participant’s use/interpretation of the sentence context. Previous studies examining children’s reading and ERPs have largely incorporated tasks that require participants to view individual words, which allows researchers to systematically control for any potential relationships between lexical items that are presented in close succession. Therefore it will be important for future investigations to carefully consider the type of ERP task that is being implemented when interpreting the results. Researchers should also pay close attention to the particular standardized measure that is used for assessing children’s academic performance given certain tasks likely require varying levels of input from several (relatively) independent processes in the brain. Fortunately, the identification of such processes could be made possible by measuring additional ERP components such as the N250, which is thought to reflect the point at which orthographic and phonological representations are combined into whole-word representations. Paradigms using rhyming and non-rhyming word pairs have
successfully elicited the N250 in children as young as 7 years old (Grossi et al., 2001), and more recently researchers have found that N250 amplitude is related to standardized metrics of phoneme blending, as well as children’s report card scores (Khalifian et al., in press). Given the importance of children’s language processing for promoting reading performance and future academic achievement, research should continue to explore the functional significance of specific ERP components and their potential use as biomarkers for children’s scholastic success.

**Changes in Aerobic Fitness versus Higher Aerobic Fitness Levels**

As techniques for measuring objective PA and sedentary time continue to improve and become more affordable (i.e., accelerometers), it will be vital for future studies to incorporate these measures to identify the potential factors that are related to changes in children’s fitness levels. Consequently, a limitation of the current study was that children’s extracurricular and habitual PA levels, as well as their time spent sedentary, were not measured; thus, the potential reasons why individual children (particularly those among the wave 1 control group) exhibited increased/decreased aerobic fitness levels remain unknown. Additionally, investigators should weigh the benefits of implementing interventions that involve multiple activity groups, as opposed to a wait-list control group. Such an approach would help to control for certain psychosocial factors (e.g., increased social interaction, improved motivation, changes in self-efficacy) that may be influenced by the program and potentially relate to cardiovascular health and/or academic outcomes.

In any case, after collapsing across waves/groups, increases in aerobic fitness demonstrated a marginal positive relationship with children’s post-test reading performance, even after controlling for IQ, N400/P600 amplitude, and other important variables. This pattern of results mirrors those of an earlier PA program that tracked changes in children’s aerobic
fitness levels and cognitive control performance over 3 years (Scudder et al., in press). Regardless of group classification, which did not differ on fitness level, increases in aerobic fitness were associated with greater improvements in accuracy for the most difficult task conditions assessing inhibitory control and working memory. As such, another current limitation was that additional cognitive control measures were not assessed, which would have permitted the examination of whether aspects of children’s cognitive control potentially mediate the relationship between increases in aerobic fitness and reading achievement. Future studies will also be necessary to determine if this association is in fact selectively related to reading achievement, or more generally related to overall academic success. Previous results from an after-school PA program in overweight children revealed a dose-response benefit of PA on math achievement, suggesting that children who spent more time in the intervention also exhibited greater improvements in math performance, yet this relationship was not observed for reading (Davis et al., 2011).

Considering children’s ERP measures and fitness levels remained relatively unchanged across time, it was not surprising to discover that the change between these variables was unrelated (suggesting no mediation). However, given that a large proportion of children in the study were below the 30th percentile for aerobic fitness, perhaps the inclusion of a greater number of moderate-to-high fit children would have altered the pattern of results. By dividing participants according to their fitness levels and matching for important demographic variables, exploratory analyses replicated previous findings by demonstrating that higher fit children had overall greater N400 amplitude and superior sentence comprehension accuracy (Scudder et al., 2014a). The current analyses, which revealed that larger N400/P600 amplitude was associated with better word recognition and spelling scores, may provide a partial account for the findings
of Scudder et al. (2014a); that is, higher fit participants not only exhibited a large biphasic N400/P600 response, but also had greater word recognition and spelling scores. Interestingly, no academic differences were observed between higher and lower fit children in the current study, which could have been due to variations in the standardized measures that were used. The inconsistency of findings across studies examining fitness/PA and academic achievement is somewhat troublesome, yet one should keep in mind that such measures are often designed in a “one-size-fits-all” fashion, and are primarily used to identify children that may be experiencing significant developmental delays. This is further exemplified by a number of fitness investigations that have incorporated behavioral measures to target specific aspects of cognitive control, with findings demonstrating that higher aerobic fitness levels and increases in children’s fitness are selectively related to superior performance for the most cognitively demanding task conditions (Chaddock et al., 2012; Hillman et al., 2014; Kamijo et al., 2011; Pontifex et al., 2011; Scudder et al. in press; Voss et al., 2011). Thus, one recommendation for future studies is the inclusion of standardized academic achievement tests, as well as cognitive control measures, to further dissociate the observed relationships between aerobic fitness and each outcome variable.

Although N400 latency was not related to children’s academic achievement, it was interesting that lower fit children demonstrated shorter N400 latency compared to higher fit children, which was an opposite pattern of previous findings (Scudder et al., 2014a). One potential reason for this discrepancy may have been the method used for quantifying N400 latency. Whereas peak latency was used in the previous study, 50% fractional area latency was chosen as a more conservative and stable approach for measuring children’s N400 latency. However, a potential limitation of this method is that within a given time window (i.e., 300-500
ms), shorter latencies may be biased towards individuals with smaller N400 amplitude, due to overall lower (50%) mean area. Regardless, analyses clearly indicated that N400 and P600 amplitude were the primary component characteristics linked to better academic achievement (even after controlling for IQ). As previously noted, ERPs in the current study and average fitness levels demonstrated little change over the course of a year, which highlights the need for additional longitudinal investigations. As it stands, the findings suggest that both aerobic fitness and the particular ERP components measured herein may serve as independent biomarkers related to children’s academic achievement.

**Future Directions**

Given the limitations highlighted throughout the previous paragraphs, there are several important areas for future studies to improve upon. To begin, objective measurement of children’s PA and sedentary time will help researchers further understand the potential reasons as to why children exhibit increases/decreases in aerobic fitness levels. If higher fitness levels are in fact related to improved academic outcomes, it will be vital for future investigations to uncover the specific participant characteristics and PA program strategies that result in the greatest fitness gains. Therefore, testing multiple intervention strategies simultaneously, as opposed to a wait-list control group, would yield far greater information concerning effective techniques for improving aerobic fitness and/or academic performance. Such an approach would also benefit from the measurement of different psychosocial variables (e.g., motivation, self-efficacy) that may influence academic or cardiovascular health outcomes. By gaining a greater understanding of the individual participants that respond favorably to a particular program, future interventions can be tailored in terms of their intensity, activities/content provided, and the time of day they occur (during vs. after-school), to meet the specific needs of children.
For example, one basic but crucial factor to consider is how school-based interventions and after-school programs differentially impact individual children. While after-school programs might afford certain children with increased opportunities for PA, future research should determine how such involvement might influence children’s otherwise normal daily routine (hence, the limitation of a wait-list control group). That is, children may have access to other means of PA, whether that is participation in leisure time or organized sports activities, in a participant- or adult-controlled setting. Similarly, implementing PA breaks in school classrooms have shown some promise for increasing children’s PA levels and academic performance (Donnelly et al., 2009), yet recent findings have suggested that such interventions are increasingly difficult to deliver due to evolving classrooms (e.g., increased technology, multiple teachers) and increased emphasis on standardized academic achievement test preparation (Donnelly et al., in review). Therefore, there is still much to be learned from designing and implementing new strategies that target increases in children’s PA and aerobic fitness levels. These attempts will undoubtedly help advance current methods to help ensure that children are receiving the most promising opportunities targeted towards their overall health and well-being.

Lastly, as investigations continue to explore the relationship between children’s cardiovascular health and academic achievement, a number of outcome measures have proven to be useful for examining the underlying cognitive processes responsible for superior academic performance. In addition to behavioral measures on cognitive control tasks, recent studies (including the current findings) have demonstrated that particular ERP component characteristics may serve as a viable index of academic success (Hillman et al., 2012; Khalifian et al., in press). Due to the lack of evidence relating changes in ERP measures to specific health/academic outcomes, future research should attempt to follow children over longer periods of development.
to determine whether modulations of these variables are associated with one another, especially
given the relative stability of children’s aerobic fitness and language-related ERP components
over the course of a year. Continuing to examine additional ERP components may also yield
valuable information concerning their role as a potential biomarker of academic success, as
demonstrated by Khalifian et al. (in press) who reported an independent association between
N250 amplitude and children’s academic marks on their report cards. Thus, ERPs will continue
to be a valuable technique for uncovering further details related to children’s cardiovascular
health and academic performance.

Conclusion

The current findings provide further evidence that PA interventions in children may be
beneficial for attenuating increases in body weight and maintaining body composition; however,
future research should continue to develop and implement effective strategies that result in
significant cardiovascular health adaptations, including improvements in aerobic fitness. Despite
no overall fitness differences between children in the intervention and control groups, analyses
accounting for individual study waves revealed a significant increase in aerobic fitness among
the wave 1 control group. Interestingly, children in wave 1 demonstrated greater improvements
for several academic tests compared to their peers in other waves, with evidence further
suggesting that the wave 1 control group exhibited larger improvements for academic composite
scores compared to the intervention group. Regression analyses also revealed a marginal
association between increases in aerobic fitness percentile and greater improvements on
standardized tests of reading. This effect was not mediated by the inclusion of children’s N400
and P600 amplitude, which were both independently related to academic performance involving
language-based abilities (children in wave 1 also had larger N400 amplitude). Thus, processes
associated with children’s access to a richer mental lexicon and greater word-related knowledge, as well as their ability to allocate sufficient resources for engaging processes that govern the identification and analysis of syntactic errors, may serve as valuable indexes of superior language-based academic performance. Fitness group comparisons replicated previous findings that higher fit children (≥ 70th percentile) exhibited greater N400 amplitude and superior sentence comprehension accuracy compared to their lower fit (≤ 30th percentile) peers, yet additional research following children over extended periods of time will be required to determine whether modulations in children’s neuroelectric indices are in fact related to aspects of children’s cardiovascular health or academic performance. Although the inclusion of several outcome measures such as ERPs, standardized academic tests, and cognitive control tasks provide a wealth of data concerning children’s behavioral performance and brain function, future studies examining children’s fitness/PA should not underestimate the importance of designing and implementing effective strategies that lead to successful cardiovascular adaptations. Accordingly, researchers are encouraged to continue exploring possible methods for influencing children’s PA levels, involving both school-based and after-school approaches. By further determining which type of approach works best for a particular child, important recommendations can be passed onto parents, schools, and important officials who all have an important voice concerning the future health of our nation’s children.
REFERENCES


in a school-based exercise program: A randomized, controlled study. *Archives of Pediatrics & Adolescent Medicine, 159*(10), 963-968. doi:10.1001/archpedi.159.10.963


hippocampal parvalbumin expression. *Brain and Development, 32*(2), 137-142.


Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38*(11), 1531-1549. doi: 10.1016/S0028-3932(00)00053-1


body mass index and academic performance in low-income elementary school children.


Mokgothu, C., & Gallagher, J. (2010). Effects of aerobic fitness on attention, memory and
decision-making in children. *International Journal of Body Composition Research, 8,*
S37-44.


electrophysiological exploration in both languages of Spanish-English bilinguals.
*Cognitive Brain Research, 22*(2), 205-220. doi: 10.1016/j.cogbrainres.2004.08.010

neighborhood density effects: What is shared, what is different? *Psychophysiology, 47*(3),
455-466. doi: 10.1111/j.1469-8986.2009.00960.x

Hoeft, F. (2014). White matter morphometric changes uniquely predict children’s reading

the development of cognitive functions during childhood. *Journal of Cognitive Neuroscience, 16*(7),
1227-1233. doi:10.1162/0898929041454451


sentence processing classes: Evidence from event-related brain potentials. *Journal of
Cognitive Neuroscience, 3*(2), 151-165. doi: 10.1162/jocn.1991.3.2.151


Schaeffer, D. J., Krafft, C. E., Schwarz, N. F., Chi, L., Rodrigue, A. L., Pierce, J. E., . . .

doi:10.1111/psyp.12227


Exercising during learning improves vocabulary acquisition: Behavioral and ERP evidence. *Neuroscience Letters, 482*(1), 40-44. doi:10.1016/j.neulet.2010.06.089


doi: 10.1097/PSY.0b013e3181d14633


