

---

# ADHD Bike Program Final Report

Wilson/Kennedy 2013

---



learn • assess • perform

***SPECIALIZED***®



## A. Executive Summary

### **Rationale**

Specialized Bikes and RTSG created a biking program at two Massachusetts schools in the fall of 2012. Students at these schools biked for half an hour before school five days a week for the period of a month.

Biking is a form of cardiovascular exercise, and has been shown to lead to a large number of desirable changes in those who participate in it. Some physical changes that can occur after exercise are well established; people demonstrate lower morbidity and mortality rates, and become more cardiovascularly fit. More recent research has shown additional affective/emotional improvements after exercise, as well as academic and cognitive improvements. These benefits of exercise can be used in schools to improve the ease and degree to which students are able to learn information and get along with one another. Since some of the cognitive improvements after exercise have to do with attention, exercise may have specific clinical effects on populations that have difficulties in this domain: for example, ADHD individuals.

Evaluation of the Specialized Bikes program was focused on changes in participants after biking. A number of different ADHD groupings were examined, as were the acute effects of a single biking episode and the long term effects of the biking program, and any physical, emotional, and cognitive changes that may have occurred after biking. Changes in brain activity were also recorded by EEG.

### **Key Findings**

#### Program (Qualitative) Results

The program was a success in the minds of the teachers, parents, and student participants. For example, 93% of students would have told a friend to join the biking program, and 91% of students described the program as fun, with 0% of students describing the program as not fun. The only negative aspect of the results was a low return rate for questionnaires sent to the parents of Wilson school students.

#### Physical (Fitness) Results

Waist circumference and BMI decreased significantly after exercise, indicating a decrease in obesity. Pacer scores improved. Balance scores improved after biking, which may be an exercise effect specific to biking. The overall results indicated that the biking program was successful in getting the students to exercise, and that the exercise led to lower levels of obesity, improved cardiovascular health, and improved balance.

#### Acute Results

A single biking session significantly improved measures of executive attention in ADHD subjects. Specifically, ADHD subjects were more accurate and faster on the Simon test after a single episode of

exercise than if no exercise had occurred. Accuracy on the TrailMaking test, a measure of attention and executive function, also increased after exercise. These cognitive improvements demonstrate that even when people are exercising regularly, an acute exercise session can lead to increases in cognitive performance above and beyond any long term improvements in performance.

### Long Term Results

There was an increase in positive mood after biking. There was an increase in the participant's ability to understand their own feelings after biking for all ADHD groups, and most non-ADHD groups. Behavioral measures of attention, as measured by the Connor's screening test, increased after biking. PEBL Connor's CPT results changed after biking, but trends could not be linked explicitly with biking. Participants demonstrated improved executive attention at the end of the biking program, with fewer errors made during biking in comparison to measurement points before or significantly after the biking intervention in the Go-No Go task. Subjects also demonstrated longer response times after exercise for a number of cognitive tests. These longer response times can be interpreted as a decrease in impulsivity, but other interpretations are possible.

### ERP Results

ERP components demonstrated a trend towards normal patterns of activity after biking in comparison to before biking, where they had smaller amplitude, and often had multiple peaks. Biking may lead to changes in the brain in students who have attention problems that cause the brain to perform an inhibitory control task more similarly to people who do not have attention problems.

### Limitations on Findings:

- Exercise outside of biking decreased when the biking program was implemented. Biking was for two and a half hours a week, and self-reported exercise decreased by roughly this amount during biking. The study may not have evaluated the effect of **additional** biking exercise as was planned because of this result, but instead measured the effect of a different form, time, and intensity of exercise.
- Subjects at the two schools were different in terms of ADHD diagnosis percentages, in terms of non-ADHD related variables, and in terms of outcomes. It is difficult to determine whether differences in program participation, differences in diagnostic category, or other differences are responsible for some of the results.

### **Potential Next Steps**

- Greater participant numbers would be needed to tease out any potential differences of the effect of exercise between 1. formally diagnosed subjects with ADHD that are taking medication, 2. that are not taking medication, and 3. that have attention issues, but are not formally diagnosed. Also, the effectiveness of medication may decrease over time, and different kinds of medication are available, so multiple medication groups may be necessary to understand the effect of exercise on the different ADHD groups that occur in the real world.

- ADHD participant proportions should be examined – biking groups may be optimal when ADHD participants are a minority in a larger biking group
- Biking specific benefits should be further examined, such as potential mood effects in more reasonable weather, increased self-efficacy in learning a skill for novice bikers, a higher exercise retention rate, etc.
- Longer duration biking should be measured, to examine the sustainability of biking, and long term benefits the current study may have missed.

## B. Introduction

### Literature Review

Exercise has a well-established positive effect on physical health. In recent years, a large number of empirical studies have examined the effect of exercise in other domains: improved thinking (cognition), greater happiness (affect), decreased severity and incidence of depression and other mental illnesses (clinical), and improved performance in school (academic). Research has sometimes found that exercise will lead to improvements in these areas. These empirical results will be introduced below in order to explain the rationale behind the Specialized Bikes ADHD intervention. There is also a movement towards more exercise in schools that is not based on research publication and academia. This movement centers on exercise programs. Exercise programs are different entities than research programs, being less well controlled and more concerned with pragmatic difficulties of implementation, as well as defining success more according to participant satisfaction and participation. Information concerning this exercise movement is provided here as well, as the Specialized Bikes intervention is also a first known attempt at determining best practice in the implementation of an ADHD biking program in the real world.

Exercise has an effect on mood in at least two domains. Firstly, it has shown to improve the mood of normal participants immediately following exercise (Lane, Crone-Grant, & Lane 2002; Sibold & Berg, 2010). The Sibold study demonstrated positive mood effects that lasted for at least 12 hours after 20 minutes of stationary biking at a moderate intensity. Long term stable mood changes can occur either by a separate long term effect or by multiple occurrences of temporary improvements in mood that lead to long term brain changes. Previous studies have not extensively examined long term mood changes after exercise in non-clinical (non-depressed or anxious) populations. Secondly, clinical disorders including generalized anxiety disorder and depression have been shown to be treated effectively with exercise (Blumenthal, 2011 is a review). One extensive and well-controlled recent study found no effect of exercise on depression or mood, though (Chalder et al., 2012). This study did not include an exercise program but rather a “prescription” for exercise where depressed individuals were told to exercise more; such a prescription is dissimilar to a structured exercise program that is regularly attended. One possible moderator for both long term mood changes and clinical decreases in depression and anxiety is a decrease in stress. Hopkins and colleagues (2012) found that long term, but not acute, exercise led to decreased perceived stress. Finally, it should be noted that exercise outdoors and in nature, not dissimilar to the biking interventions attempted here, has been examined in at least

one study (Barton & Petty, 2010). Outdoor vs. indoor exercise comparisons were not made, but exercise near water was shown to have a greater positive effect on mood than other forms of exercise (all forms showed significant pre vs. post changes in mood).

The link between exercise and cognition has also been established (Smith et al. 2010 investigates the effects of long term programs on cognition; Chang et al. 2012 examines acute exercise and cognition). Effects tend to be small but significant, and tend to occur in executive function, memory, and attention domains. Although there have been many studies that have examined cognitive function in adults, and a number of studies of preschool kids, purely cognitive studies of exercise & cognition in children are more rare (but they do exist: for example, Best (2010) reviews the acute and chronic effects of exercise on children's executive function). There are, however, many studies of children in the context of academic performance. A review of such school studies is contained in a recent CDC white paper (CDC, 2010); exercise in school generally has a positive effect or no effect on academic and cognitive performance. School study quality tends to be poor, however. A recent, well controlled school study (Hill et al., 2010) found cognitive gains after the second week of an exercise program, and demonstrates that rigorous research can also demonstrate positive effects of exercise on cognition.

ADHD specific research involving exercise is rare. Archer & Kostrzewa (2012) suggests that ADHD may be improved by exercise and notes that ADHD individuals tend to have deficits in many areas that are improved by exercise (see Halperin & Healey (2011) for another similar kind of review). Specifically, ADHD children tend to exhibit lower levels of emotional stability and higher levels of negative affect, among other mood related traits (De Pauw and Mervielde 2011). Preliminary evidence suggests that ADHD children may also demonstrate improved executive function after a long term exercise program (Gapin et al. 2011). Gapin and Etnier (2010) found a link between longer durations of daily physical activity and executive function performance as well. A recent and well-controlled study (Pontifex et al. 2013) has demonstrated the effect of a single bout of exercise on the Flanker test, a test of executive attention. Research on long term effects of exercise on attention and mood is lacking in the ADHD population; since these are significant areas of concern for this clinical group, such research would potentially be very useful.

Besides being a significant area of research, exercise programs in schools are a significant movement. Naperville Central High School in Illinois has a well-publicized physical activity program that implements exercise throughout the day. "Zero hour PE", an hour of activity at the start of the day, was used as a tool for subsequent academic improvement, and served as a model for many subsequent school programs. The program was covered and publicized in the book "Spark" by Dr. John Ratey, who has advocated for exercise based-PE throughout the US and around the world. Gains at these kinds of programs are anecdotal, and tend to use the following logic: 1) In previous years/time periods, problems have occurred at the school in an area(s) 2) Following exercise, these problems are alleviated according to the judgment of teachers, the self-report of students, and sometimes basic calculations of percent changes on a few measures 3) These changes must be due to the exercise, because such gains were not reported in previous years. This kind of program evaluation is not suitable for research publication, but

can be sufficient to convince school administrators to support programs, especially when program results are supported by parents.

These kinds of program results should not be completely discounted, though, because exercise programs are not identical. In laboratory settings, exercise can be controlled precisely, and recruitment and drop-out issues are navigated successfully (or the research would not be published). Real world exercise programs can vary widely in quality and have real world issues that do not exist in the laboratory. Best practices for the creation of school exercise programs are difficult to determine completely “objectively” since achieving proper controls can be very difficult; practices that are associated with success tend to be emphasized, and those with failure tend to be dropped. It is important, therefore, when testing a real-world exercise program, to not only collect objective, quantitative measures, but to also get a qualitative report of the program that was run. This evaluation allows an analysis of variables that are not specifically measured in the program, and can lead to the formation of hypotheses to be formally tested later, and in the meantime become part of tentative best practice in program creation. It is better to give program creators an idea of what has and has not worked in the past than to give them no advice whatsoever.

## C. Evaluation

### Research Questions

1. Will students display positive changes after biking exercise? Will mood, attention, peer liking, cardiovascular health, and behavior improve post exercise in comparison to pre exercise? Will obesity measures decrease?
2. Will there be differences in improvements for ADHD students in comparison to non-ADHD students after biking, especially in regards to attention? Will the opposite relation hold?
3. Are there correlations between mood and attention gains on the one hand, and exercise intensity and exercise frequency on the other?
4. Are there differences in exercise results according to how attention issues are diagnosed? Does a strict ADHD diagnosis category mean that students perform more poorly, or does the potential effect of medication mean that undiagnosed students with attention difficulties demonstrate worse performance? Do groups demonstrate different performance levels on cognitive and affective variables after biking?
5. If there are differences between groups after biking, to what degree can we attribute any changes to the biking intervention itself?
6. What brain changes occur (as measured by QEEG and ERP) after exercise in ADHD subjects? Can improvements in QEEG be seen at the group level and by subtype groups? Does the P3b ERP wave change significantly after biking, and does the N2 wave change after biking on inhibitory tasks?

### Program Questions

1. Will students maintain levels of outside exercise while participating in a biking program?

2. Will students maintain current levels of sleep while participating in a biking program in the morning?
3. How difficult is it to get middle school students to bike frequently? Are there specific difficulties with ADHD students?
4. How do parents and students respond to participating in a program that is labeled as an ADHD program?

## D. Methods

### Participants

Students at two separate Middle Schools in were recruited to participate in a 30 minute, five day a week before school bicycling program for one month in a school district in Eastern Massachusetts. In one school (Wilson), students were asked to participate in a cycling program on the basis of having an individualized educational plan (IEP) that targeted attention difficulties, that is, students with or without an ADHD diagnosis who were flagged internally in the school as needing support for their attention difficulties. In the other school, (Kennedy), students were first recruited on the basis of having an IEP targeted to attention difficulties; however, study participation was open to all students in the seventh grade.

### Measures

#### i. Questionnaire

A questionnaire was handed out to students during measurement periods (see Appendix B). The questionnaire was meant to address some interesting questions, and examine a number of potential confounding variables. The questionnaire asked subjects to rate their level of exercise enjoyment, and to rate their liking of their peers, both on a five point Likert scale. In previous programs, RTSG has found that peers that exercise together tend to like each other more, but this variable has not been examined in the context of biking, or in the context of an intervention with ADHD individuals. Another variable that was examined was the amount of time spent in front of a screen time at various points of the day. Time spent watching a screen has been linked to higher morbidity and mortality rates, but that was not the reason for its inclusion in the survey. Two questions were meant to be addressed: do ADHD individuals spend more time in front of screens than non-ADHD individuals, and if screen time before school is decreased (due to exercise) is this screen time made up at other times outside of school?

The other purpose of the questionnaire was to look for the presence of potential confounding variables. There are a number of ways that results from an exercise program could be attenuated. Since biking was early in the morning, participants may get less sleep during the program than they did before it. On the questionnaire, we asked subjects how much sleep they had during the intervention, and before it, both on week-days during the intervention, and on week-ends (to see if they were making up sleep at this time). We also asked participants how awake they were in the mornings during biking (a 5 point Likert scale, as another way of addressing this question. Participants may also have performed

less other physical activity during a biking intervention than they did before the intervention, so the hours of outside physical activity were also recorded. If either the amount of sleep or outside exercise goes down during the intervention, the effect of biking may be counterbalanced by an opposite effect due to lack of sleep or lack of outside exercise.

## ii. Qualitative

The student participants, their teachers and parents were asked their opinions of the Specialized cycling program at the end. Surveys were distributed upon completion of the program at each school. The feedback that was solicited was used to try and determine practical strengths and weaknesses of the cycling program, to determine how cycling could better be implemented in schools in the future, and to determine aspects of the current program that were strong and should be included in future programs.

## iii. Fitness

### Pacer

The Pacer, or beep, test is a measure of cardiovascular fitness. Subjects run between two points that are spaced twenty meters apart. A recording that plays beeps at regular intervals is played during this run; subjects run the distance in the time between beeps. The duration between beeps stays constant for a number of beeps and then speeds up, and each period of constant duration is considered a level. Subjects are scored according to the level they are on when they fail to run the interval in the allotted time between beeps.

Pacer scores contain both a level indicator, plus a number of successfully completed shuttle runs on that level. Pacer scores were converted into decimal numbers in statistical analysis by dividing the number of shuttle runs made on a certain level by the total number available on that level. On the pacer, the number of successful shuttles that need to be run increases as the level increases, so for level 4 on the pacer, a subject needs to run 9 shuttles to advance to the next level, whereas for level 15, a subject needs to run 13 shuttles. A subject who was on level 4 and was able to complete 5 shuttles on that level was given a score of  $4(\text{level}) + 5(\text{shuttles run on that level}) / 9(\text{shuttles needed to be run to get to level 6}) = 4.556$ , for example. Pacer scores were hypothesized to improve as a result of biking, due to an increase in cardiovascular fitness.

### BMI & Waist Circumference

BMI is a traditional measure of obesity, whereas evidence suggests that measures involving waist circumference have more predictive value for future health problems. Both measures were used in the current study. BMI's and Waist Circumferences were hypothesized to decrease as a result of biking.

### Balance

Balance was tested by using the Equilibrate system made by the company Balance Engineering. Subject's stood on force plates, and observed a point on the wall in front of them approximately 2 feet away. They wore a vest with diodes on standard anatomical markers. Two cameras captured postural sway data and force plates captured center of gravity/shift data. Subjects were assessed with their eyes open and closed, and standing on one foot or both feet. Testing occurred in the following order: two feet eyes open for 15 sec, two feet eyes closed for 15 sec, right foot eyes open for 10 sec, left foot eyes open for 10 sec, right foot eyes closed for 5 sec, and then left foot eyes closed for 5 sec. It was hypothesized that balance would improve in the eyes open conditions with regular biking.

**iv. Affective**

Child PANAS

Mood was measured using the child version of the PANAS (Laurent et al., 1999). The PANAS was originally developed by Clark and Watson (1991) as part of a tripartite model of anxiety and depression. Their theory was that anxiety was characterized by negative emotions, whereas depression was characterized by both negative emotions and a lack of positive emotions. The PANAS scale was used to assess emotions in regards to their positivity and negativity for the purposes of studying and diagnosing anxious and depressed individuals, and it was later used in non-clinical populations. The child PANAS was adapted from the regular PANAS scale and consists of 27 items, 15 of which belong to a negative emotion subscale, and 12 of which belong to a positive emotion subscale. Note that three additional items were included on the PANAS given to students; these items were not included in the evaluation, and were put in for reasons unrelated to the intervention. It was hypothesized that after biking, students would experience more positive emotions, and fewer negative emotions, as measured by the PANAS.

Trait Meta Mood Scale

Besides a current indicator of mood like the PANAS, some indication of changes in emotional intelligence was desired, and the Trait Meta-Mood scale was used to assess any changes that might have occurred as a result of biking. The Trait Meta-Mood scale measures three components of attitudes or beliefs people have towards their own emotions: the ability or degree to which they attend to their feelings, the clarity of their feelings (i.e. the degree to which they can interpret those feelings), and the degree to which they can repair or deal with their feelings. In the current study, the Trail Meta-Mood Scale for Children (TMMS-C) was used along with the Revised Life Orientation Test (Rockhill & Greener, 1999). This resulted in a 22 item scale where subjects give responses on a 5 point Likert scale. Response items are associated with the various aforementioned attention to feelings, clarity of feelings, and mood repair subscales as well as a subscale for dispositional optimism, which measures how positive an outlook on life the child has. Tentative hypotheses were that all of these measures would improve after biking.

**v. Cognitive**

PEBL Go/No Go and PEBL Connor's CPT

Continuous performance tests (CPTs) were first developed by Rosvold et al. in 1956. The Connor's version of the test involves the presentation of a number of different letters on the screen consecutively, and the pressing of a button when any stimulus other than an "X" is presented. The Connor's Continuous Performance test measures two components associated with attention: sustained attention and response suppression. The test takes 15 minutes, and the PEBL version is done on a computer. The PEBL version of the Go/No Go is similar to the Connor's CPT. It involves two conditions, one where subjects press a key when a "P" is presented in one of 4 quadrants, and one where subjects press a key when an "R" is presented. "P" and "R" are the only two stimuli presented, and "P"s are presented more often. The test is done in two 5 minute blocks, with a change of instruction at the half way mark such that in the second half "P" is the cue stimulus, and so measures sustained attention less (since subjects need only concentrate for 5 minutes at once in comparison to 15 for the Connor's), and can thus be thought of as a "purer" measure of response suppression, since differences in scores are more probably a function of this process.

### PEBL Flanker

The Flanker task (Erikson & Erikson, 1974) is a relatively simple procedure where an arrow stimulus is presented in the center of a computer screen. The arrow can point to the left or the right. The subject's task is to press the left <shift> key when the central arrow points left, and to press the right <shift> key when the central arrow points right. In the PEBL flanker, there are three conditions in total: a congruent condition, a neutral condition, and an incongruent condition. In the congruent condition, arrows appear on the left and the right of the target arrow, and point in the same direction. In the incongruent condition, arrows appear on both sides of the target central arrow, but point in the opposite direction. In the neutral condition, no additional arrows appear. In the PEBL Flanker, 12 practice trials are performed, and then 120 actual trials, split evenly between the three congruency conditions, and with an even split in the direction of the central arrow, and thus the direction of the expected response. The Flanker task is linked with theories of attention. More specifically, it has been linked to a component of executive attention labeled interference control (Mulane et al., 2009). This construct is measured in the Flanker by examining subject behavior according to congruency of stimulus, and also by examining subject responses on an item basis. In previous research, reaction time (RT) following an incorrect response (Kerns et al., 2004) and for correct trials following a match trial (Mathewson et al. 2005) was measured; changes in these measures are thought to indicate top down control by executive attention. Note that median RTs are often used instead of mean RT on this test for ADHD individuals, because this measure is thought to better reflect central tendency in this population (Pontifex et al., 2012).

### Connor's Screening Test

The Connor's ADHD Test, written by C. Keith Conners, Ph.D., involves a series of questions concerning behaviors that are associated with attention. The test is meant to be given to parents of children who may have ADHD, and can be used to indicate quickly if a child's clinical symptoms require more investigation and a possible diagnosis of ADHD. The test was used in the current intervention to give a more global, behavioral measure of attention. Tests like the Flanker and the Simon measure

specific aspects of attention; the Connor's can be used as an estimate of overall attention. The test was slightly modified and also given to the students themselves as a measure of self-reported attention function. So the test has potential reporting and self-reporting biases and inaccuracies, but describes real world attention well.

**vi. Acute Specific Tests**

PEBL Simon Test

In the Simon test, subjects are presented with an item in relation to some central point of fixation (to the left or right of this point). Subjects then have to make a response to this item with either their left or right hand. In the PEBL version of the Simon, colored circles are presented on a computer screen. The circle is presented either directly in the center, or some distance to the left or to the right of center. If the circle is red, subjects are instructed to press the left <SHIFT> key, and if the circle is blue, the right <SHIFT> key. The Simon test is similar to the Flanker in that it is considered to be a test of executive attention/interference control (Mulane et al., 2009). The interference occurs when the side of the screen that the circle is presented on is opposite to the side of the response. For example, a red circle is presented on the right side of the screen; subjects should press the left <shift> key with the left hand, which is the side opposite to the one where the circle was presented. Generally, it is expected that subjects will take longer to perform correctly for incongruent items (different location and response sides) in comparison to congruent item, and that this time difference is the marker of an interference control process.

The Simon test is substituted for the Flanker task for the acute measures. The Flanker task has already been tested after exercise in an acute population (Pontifex et al., 2012). The acute measures in our intervention measure the effect of a single bout of exercise on top of the long term effects, while the Pontifex study measured the effect of a single bout of exercise in isolation, it is true. However, if no long term effects of exercise were found for the Flanker, any positive acute results would duplicate the Pontifex result. Substituting the Flanker test, which measures the same theoretical construct, thus guarantees a novel result, since it has not been studied with an ADHD population during a long term exercise program.

Five Point, FAS, and Animal Naming

The FAS and animal naming tasks are both verbal fluency tests that involve subjects producing words specified by certain conditions. Subjects must generate as many words as possible in a three minute period. In the case of the FAS, the words are generated according to the first letter in the word; it must start with an F, then an A on a second trial, then an S on the last trial. In the animal naming task, generated words must belong to the category of animals. Both tests thus measure similar things, and involve executive functions in creating a strategic search process to generate new words, and the cognitive flexibility to switch from one strategy or category to another (Troyer et al., 1997). However, the FAS has a phonetic component and the animal naming test has a semantic component, that the other test does not have. The five point test is a non-verbal analogue to word fluency tasks. Subjects are given a sheet of paper with 40 five-dot matrices, and are asked to produce as many different figures

as possible within 3 minutes. Note that the five point test does include a mandatory, unpracticed motor component, whereas verbal tests include spoken answers, or written answers, both of which involve well practiced and more complex motor movements.

### Symbol Digit

The Symbol Digit Modalities test was developed by Aaron Smith originally as a screening measure for cognitive dysfunction in children or adults. In this test, subjects are presented with a key consisting of abstract symbols and numbers associated with those symbols. There are rows of the abstract symbols on a page, and the subjects must, for each symbol, retrieve the associated number from the key (or memory), and write it under each symbol. The number of correct responses made in 90 seconds was recorded. The task is similar to the fluency tests in that subjects must produce responses that are restricted by a cue, but instead of generating as many unique responses as possible, subjects must instead produce the one correct response from a key for each item. Thus, while the fluency tests involve executive functions, the symbol digit involves facets of attention, including divided attention, complex visual scanning and tracking (as well as other non-attention processes like perceptual speed, motor speed, and memory).

### TrailMaking Test

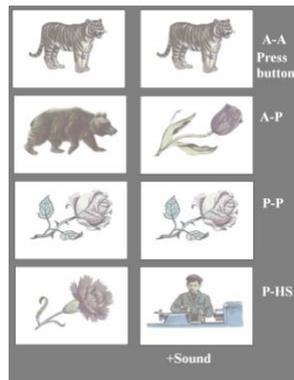
The TrailMaking test was originally constructed as Partington's Pathways (Partington & Leiter, 1949), but has since been incorporated into a number of batteries and has a long history of use in assessment. The test requires subjects to connect circles in a specified order. In the TrailMaking A condition, circles are connected in increasing numerical order. Subjects start at the circle labeled with a one, must draw a line to the circle with a two, and continue this process until circle 25. In the TrailMaking B condition, subjects alternate between letters and numbers, and so connect circle 1 to circle A to circle 2 to circle B etc... The test has been used to measure divided attention, and visual scanning and tracking much like the Symbol Digit test. It has also been used to measure executive function. The TrailMaking B test involves an additional process of switching between letters and numbers in comparison to the TrailMaking A test, and so different mathematical equations have been used to analyze the difference between the B and A tests, which is thought to indicate the function of this specific switching process.

### **vii. QEEG and ERP**

EEG was recorded by means of the Mitsar (Mitsar, Ltd.) amplifier as manufactured by EVOKE Neuroscience Company from 19 electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2 sites in the International 10-20 system) with 500 Hz sampling rate in 0.3 – 70 Hz frequency range in the following conditions: 1) eyes opened (EO) – 3 minutes, 2) eyes closed (EC) –3 minutes, and 3) a modification of Go/No Go task (20 minutes). Low cut band pass filters were set at 0.3 and high cut band pass filters were set at 50 Hz. All data was processed in WinEEG, using independent component analysis (ICA) to detect and remove eye movement artifacts. Movement artifacts were identified and artifacted from the data. Fast Fourier transforms were used to produce spectral analyses of eyes open

and eyes closed data. EEG spectra and ERP components were compared to an age matched normative database (HBI).

The Go/No Go task consisted of 400 trials presenting visual stimuli to participants every 3 seconds. In the task four categories of stimuli were presented: 1) 20 different images of animals – referred to later as **A**, 2) 20 different images of plants - **P**, 3) 20 different images of humans presented together with an artificial sound – **HS**. (Each sound was composed of four pure tones of 500, 800, 1100, 1400 Hz of 20 ms duration). Trials consisted of presentation of a pair of stimuli with interstimulus interval of 1.1 sec. Four categories of trials were possible: **A-A**, **A-P**, **P-P**, and **P-HS**. The trials were grouped into four sessions with one hundred trials each. In each session a unique set of five **A** stimuli, five **P** and five **HS** stimuli was selected. Each session consisted of a pseudo-random presentation of 100 pairs of stimuli with equal probability for each category and each stimulus. Participants were instructed to press the “Go” pushbutton as fast as they could when they saw two animals presented in a row. If they saw an animal, then any other image, they were not to press the button.



The data were stored on the hard disk in the linked ears reference montage and processed offline by means of WinEEG software.

While EEG provides insight into cortical self-regulation, ERP gives a view into how information is processed within the brain in various components, generally divided into pre conscious and conscious awareness. It has been documented elsewhere in the literature that children with ADHD have different neuroelectric activity than their ‘normal’ peers. Specifically, four subtypes of ADHD have been identified in the work of Kropotov and colleagues. These EEG spectral subtypes include (1) elevated theta/beta power ratio, (2) theta at the frontal midline, (3) excess alpha at three potential sites: parietally, occipitally or along the sensorimotor strip (in the form of the mu rhythm) and (4) excess beta at the frontal midline. While North American publications focus heavily on the prevalence of the elevated theta/beta power ratio subtype of ADHD, publications in Europe, Russia and Australia provide a more complex view of ADHD at the cortical level. Biker’s ERPs are analyzed in terms of these subtypes.

While there are many ERP components (components are defined in terms of their onset, peak, and polarity), for the purpose of the present study, four components are focused upon. The first component (P3b stm 1) is relevant to the presentation of the first stimuli during go tasks. Specifically, if an animal was on the screen during the continuous performance task, the participant was instructed to

wait to see if another animal came next in the sequence, and if so, press the push button as quickly as possible. P3b in this paradigm of stimulus presentation is related to the allocation of attention resources. When the P3b is well defined for the Go tasks, the participant consistently will see the first animal on the screen and await a possible second animal. The P3b should be larger for Go tasks than No Go tasks during stimulus one. In No Go tasks, the participant should quickly realize that the image is not of an animal and relax their state of attention until the presentation of the next pair of stimuli.

The presentation of the second stimulus, when it is an animal and the Go response of hitting the pushbutton is required, elicits a P3b wave again. The P3b wave here provides information about the participant's allocation of attention during the presentation of the second stimulus. The greater the amplitude of the wave, the more attention resources are allocated, and the shorter the latency of the wave and the faster the information processing in the brain. When the first stimulus is an animal, and the second is not, two No Go components are relevant. The first occurs over the central motor cortex, and the second occurs over the parietal cortex. The N2NoGo appears when there is a non-target stimulus. In this case, the participant is preparing for the second animal to appear, and when it does not, the deviation creates a negative component, called N2. The P3NoGo over the parietal cortex is associated with the participant engaging attention resources and inhibiting his behavioral response of hitting the push button.

Changes in P3b and N2NoGo waves were examined in ADHD and non-ADHD subjects after biking, and comparisons to norms were made.

(Acknowledgement is provided to Jay Gunkleman, Yuri Kropotov and Elena Labkovsky for their expertise during consultation for data processing and interpretation)

## **Procedure**

At both Wilson and Kennedy schools, an hour lecture discussing the program design, and possible benefits and risks of bike riding was offered to perspective parents prior to the start of the bike program. The principals at both schools and the superintendent of the district reviewed and approved the proposed study design and all data collection materials. Teachers were recruited to supervise morning rides based on their experience riding bikes and with students with attention difficulties. Teachers determined the best routes to take each day based on weather and traffic conditions with the objective of maximizing safety and the duration of continuous riding. Prior to the start of biking, all students were provided with a brief overview of road safety, turning signals, and so forth. Students received professional bike and helmet fitting on the morning of the first ride.

A cross over design was used to best utilize the thirty bikes and maximize program participation. All participants were informed that attending the majority of bike sessions would enter them into a raffle to win a bike or to win a gift certificate to the local mall or iTunes. From mid-September through mid-October, the bike program was conducted at Wilson. From mid-October through mid-November, the bike program was conducted at Kennedy. Wilson had premeasures taken prior to the first biking session, post measures conducted in mid-October, and a retention measure in mid-November. Kennedy had baseline measures taken in mid/late September, pre measures taken in mid-October and post

measures taken in mid-November. Acute measures were taken at the mid-point of each biking program. For Wilson, this midpoint was at the beginning of October, and for Kennedy, the midpoint was at the beginning of November

All measurement points (base, pre, post, retention) consisted of the same data collection methods. Cognitive, mood, balance and physical measurements were taken for all students on the same day. A smaller subset of students were selected for neuroelectric measurements based on reported attention difficulties, either a pre-existing ADHD diagnosis or an elevated Conner's screen score of 10 or more of a possible total score of 30, and no reported pharmaceutical treatment of their attention deficits. Twenty five students were selected for neuroelectric assessments, one was excluded after it was evident that an unreported tic disorder created excessive noise in the electroencephalographic (EEG) signal, and two were absent from school during the post measure data collection period. Twenty two remained for analysis.

## E. Results

### Qualitative Results

#### DID YOU HAVE FUN?

A total of 54 students at both schools completed the exercise program. 49 students reported the program as "FUN", with 5 saying they had fun "SOMETIMES". No students reported that the program was "NOT FUN".

Kennedy Middle School students were very positive about the cycling. Of the 29 participants, 28 reported that the Specialized Cycling program was "FUN" and 1 reported "FUN SOMETIMES". At Wilson Middle School, 21 students reported it was "FUN", with 4 answering it was "FUN SOMETIMES".

No student at either school said the program was not "FUN", a testament to the enjoyment cycling brings, but also an indication of the teachers' leadership in creating a positive/inclusive school cycling community. Having fun and beginning the day with enjoyment is important for students. Mood is an important component of learning, as well as developing social relationship skills. It was emphasized to teachers during recruitment that the objective of the riding program was for kids to have fun while riding fast (to keep their heart rates up) and staying safe.

#### WOULD YOU TELL A FRIEND TO JOIN?

50 of the Cycling program participants said that they would "TELL A FRIEND TO JOIN", with 2 saying "MAYBE" and 2 "NO". Kennedy students overwhelmingly said that they would "TELL A FRIEND TO JOIN" 28 to 1. At Wilson, 22 students reported they would tell a friend, with 2 saying "MAYBE" and 1 saying "NO".

We posed this question to the student Cyclers for two reasons. The first reason was to better gauge the interest and enjoyment of the program by the students. If students would tell a friend to join the program, it has positive "street cred" within the school community and is view as favorable, or

“cool”. Also, the question is designed to gauge interest in expanding the Specialized Cycling program into a sustainable, daily/yearly club that has many, many members at each school; creating a cycling “gang”. In both schools, students genuinely enjoyed their participation, were proud of their affiliation, and want to start up in the spring of 2013.

#### WOULD YOU DO IT AGAIN?

38 of the students said that they would participate in the Specialized Cycling program again, and 3 said they would do it if it was scheduled in warmer weather. 6 of the students said “MAYBE” and 7 students would not participate in the program again.

At Kennedy Middle School, 22 of the 29 reported “YES”, they would definitely participate again, with another 3 saying conditionally “YES...IF IN WARMER WEATHER”. 3 students said “MAYBE” to further participation, and 1 said “NO” and that the program was not for them.

Over at Wilson Middle School 16 students voted “YES”, with 3 in the “MAYBE” category. 6 students responded “NO”, however 5 of these students were early drop-outs of the program. 3 students cited a lack of a ride from a parent, and the other 3 did not specify the reasoning behind their thoughts.

#### WHAT DID STUDENTS “LIKE” ABOUT THE PROGRAM?

In an open-ended question, Kennedy students were asked to write-in what aspects of the program they most enjoyed. Student responses were coded and analyzed. 11 students reported that the program allowed for scheduled time for “BEING WITH FRIENDS”. Another 11 students simply said that “BIKING” was the best asset, with 1 saying “BIKING FAST”. One lone student reported that he enjoyed the brain assessments and learning from the scientists who ran the study.

At Wilson, 7 students reported “BEING WITH FRIENDS”, and 7 said it was “BIKING” that was the biggest lure for the program. The cycling course for the Wilson students had many hills, resulting in 6 students reporting that they enjoyed “SPEEDING DOWN HILLS, GOING UP HILLS” the most. Anecdotally, the scientists joined in on the riding program once per week and witnessed one student speeding down a hill – no handed – and looking over at the scientists and reporting “I am going 25 miles per hour with no hands”. Though dangerous (the behavior was corrected), this incident highlights that the kinds of development that can take place during cycling. Cycling allows for more “autonomy” and “self-determination” than normal educational or school-based programs. Students had the choice to go faster or ride more moderately. They could jump off curbs or drive cautiously on the street. It was up to the students themselves.

“Exercising in the morning” and allowing students to be ready for classes was the biggest draw for 5 of the students. Students enjoyed being mentally ready for the beginning of the day.

#### HOW WOULD STUDENTS MODIFY THE PROGRAM?

Kennedy students (13) mostly requested riding in warmer weather. 5 students would change nothing, reporting that they were thrilled with all aspects. 2 students would like the riding later in the day because it was too hard to wake up so early. 2 students would like a more challenging course and 2 did not enjoy the brain assessments. 1 student would have liked to “bike longer” and 1 would have liked to listen to music while riding.

Wilson students (8) reported that the program was too early in the morning. 4 of the students would change “NOTHING”, whereas some students would prefer differing routes each day to add novelty, and 4 others wanted routes that had “NO HILLS”. Interestingly, 3 students reported that they wanted more hills. Other responses were bike only when warm, not wearing heart rate monitors (a measurement necessary for the experimental study design), and biking every day including weekends.

Worthy of note here is that the exercise participation in the Specialized Cycling program was exemplary. No students had a problem with participating in the biking. The teachers and scientists treated the students with kindness and respect. And Specialized contributed world-class bicycles that did not fail, but rather stood up to the beating that only Middle-Schoolers can offer to a bike. Specialized should be acknowledged that their bicycles were a hit with the students, and never let them down. The only complaints were time of day and year issues...that is an indication of a program well run and one that pays attention to the needs and requests of students. The weather cooperated for the majority of the study, however there were two rain days at Wilson where students played small sided games in the gym, and there was a hurricane during the bike program at Kennedy, keeping the students out of school for two days and in the gym due to fallen branches for an additional day. The students much preferred to be on their bikes than in the gym.

In the voice of one student, “I enjoyed it...even though it was too early”. Another simply said: ““Awesome!” That is very high praise indeed. A third student shows the pride and adventurous spirit the Specialized Cycling program offered him by wanting to “Ride to center of Natick and back.” These are certainly powerful endorsements.

### Student Results

Listening and incorporating the lessons learned from the students in the Specialized program going forward are keys to achieving a higher success. No program can make everyone happy, but listening to students will help it become one that is not cookie-cutter, but specific to each individual, unique school community that utilizes it.

Drop outs were often due to non-program causes. Half of the students who came most infrequently or dropped out cited that their parents were unwilling/unable to drive them to school so early. A possible solution to this issue is offering the Specialized Cycling program to students at various times of the day so as to ensure that all students willing to participate are able to do so. A first period class of biking would be possibly be better, because it would eliminate transportation issues, and would potentially increase the amount of sleep that student’s get (we found evidence that students were feeling tired when they arrived at school, both before and after participation in the cycling program, see

below.) This is something that the district superintendent Dr. Sanchioni is considering for the 2013-14 school year.

The music-requesting student does not raise a valid concern. It is thought that listening to music may divide attention while biking, and increase the risk of an accident. Anecdotally, the student who requested the iPod music was the same student that hit a pedestrian, with no injuries to either party.

Perhaps the most important qualitative “discovery” of the Specialized Cycling program was how the students value “friendship” and being with “friends”. General peer liking results are given further below, both before and after cycling. It should be noted, though, that purely in the context of cycling, students had very good social interactions. The Specialized cycling program was egalitarian and democratic. The faster student riders encouraged, and sometimes rode with the slower student riders. Everyone got along and supported each other while biking. So regardless of any long term changes in social liking and social skills, cycling became an occasion where inclusiveness occurred and social skills were practiced.

## PARENTS

Parent surveys were sent home with students upon completion of the Specialized Cycling program at their school. In open ended questions, parents were asked to identify all of the “Strengths” and “Weaknesses” of the Specialized program and to suggest “Modifications”. The Kennedy parents returned 24 of 29 surveys, whereas the response rate at Wilson was 2 of 25. This difference in response may represent a difference in parent involvement at the two schools.

### Parent Perceived Strengths

The parents of Kennedy students identified a number of “Strengths” that help to understand why the program was so well received. The number one response by the 24 parents was that the program was a terrific vehicle to deliver regular cardiovascular exercise, which was cited 6 times. Strengths cited 5 times were that “My Child Enjoyed It”, “It Was Great”, and that the parents appreciated the program as a tool for their child for “Biking With Friends and Teachers”. An unforeseen consequence of the program was that 5 parents reported that it “Helped My Child and Family Get Ready For School” and another 4 said that “My Child Looked Forward To Coming To School In The Morning”. This underscores the importance of making school “FUN” and something to look forward to. It not only affects the student, but the whole family. Moms’ and Dads’ job gets easier to wake the child up and get them ready for school if school is fun and something to look forward to. 3 parents reported that their child was in a “Better Mood”, which is great for both parents and child alike. Other strengths mentioned once were: the opening lecture by the scientists presenting the neuroscience of how exercise affects cognition and emotional regulation, that the program was well organized, weight loss, better grades, more energy and that a student became a better bike rider.

Select quotes for the surveys highlight the strengths of the program. One mother wrote that “She enjoyed program in spite of early hour, she hates getting up early, and weather - she hates cold...this speaks volumes to me.” Another mom indicated: “Getting whole family out of house earlier

each day, he has been in a much better mood each day he bikes.” Lastly, a supportive father in sharing that “My child was excited to ride with friends in group, was more active, happy, took pride in what she was doing, better in school!”.

### Parent Perceived Weaknesses

Surprisingly, 18 parents indicated that there were no weaknesses or things they would change for the program. The parents who saw the scientists around the town and school were so complementary, and thankful for including their students...making for very rewarding work for the teachers and Scientists alike.

Of the weaknesses identified, 6 parents cited the weather and that it was so cold in the mornings. Other single responses were that it was too early in mornings, the group of riders was too big, and that the group was not riding fast enough for one student.

### Parent Suggested Modifications

Keeping with the theme of the cool/cold weather, 6 parents would like to see the program shifted to the warmer months of September and Spring when the weather is “warmer”. Other responses: get help from parents and teachers to reduce group size, keep doing it all year, or 4 of 5 days a week, ensure every student has access to program, particularly before state testing, have a longitudinal study, have no testing, & have a wrap-up presentation with funny moments from the biking (Minor Collision with a Pedestrian Incident, Deer Almost Hit)

### Summary Parent Quotes

One mother reported: “This program was wonderful and we are so glad we were able to participate in it. He definitely was less moody, more upbeat. We used to have to beg him to ride his bike but I think he realized how it made him feel good or better about himself.”

A father said that he: “...heard numerous questions and comments from co-workers and friends who live near Middle Schools about the biking kids they saw...all had positive things to say when it was explained. Overall a very positive experience for her.”

“[My son] would have liked to learn more about the science part...he was very interested in the studies...overall everything was great, [and he] was sorry when it was over, he won a bike which is very exciting, a couple of mothers and I were talking and thought it would be interesting for parents to get together and share perspectives.”

### TEACHERS

The classroom teachers for the first period after cycling were asked, via an email survey, their thoughts on the Specialized cycling program. The responses from teachers are an important component of the qualitative evidence; having teachers onboard and supportive of the Specialized cycling program

is important both for the dynamics within the school community, but also the teachers care for both the cognitive/intellectual and emotional regulation/behavior of the students.

### Teacher Perceived Strengths

The classroom teachers of the cycling students were very supportive of their student's participation. One teacher noted that the "kids (were) enthusiastic about the program, loved to talk about it - for most it gave structure to morning". The same teacher noted that for a couple of students it was difficult for them to get to class on time. In the future, scheduling should be addressed. And at this School District, the Superintendent suggested that more time might be built into the day to accommodate the extra time needed to both bike further, and get to class on time.

Another teacher noted the buzz around the school regarding the Specialized Cycling program: "students are excited to be part of program, they came into homeroom energized". This is a positive way to start the day for any teacher. Having the students alert, ready to learn, and in a positive frame of mind both help teachers teach – and help learner learn.

Lastly, and perhaps most intriguing, is that at the four week mark some students were just beginning to exhibit improved classroom behavior leading to academic performance. - Did not notice vast changes in behavior until this week...he seems much calmer. This is interesting because a parent suggested "longitudinal study with same cohort". This mother agreed with the teachers that the some of the benefits of the program were just starting to occur. Perhaps more changes in cognition an emotional regulation might be captured in both a longer program and looking at the student's progress over a number of years/decades.

### Teacher Perceived Weaknesses

Very few weaknesses were reported by the classroom teachers. One teacher noted that one student had a body odor issue, a very common occurrence in Middle Schools. The PE teachers at the school suggested that next year the program would be a nice segway to personal hygiene lessons in the health class.

Also, the grades reportedly went down significantly for one student. However, the cause of this decrease cannot be directly attributed to the cycling program.

### Teacher Suggested Modifications

The classroom teachers give a resounding "thumbs up" for the Specialized Cycling program by requesting a lunch time ride in addition to the morning program. The reasoning was that the classroom teachers of the students after lunch might also benefit from the changes in intellectual an emotional optimizing the biking seemed to contribute to cognitive/emotional benefits for after lunch classes

One classroom teacher noted that the program "Looked like fun, loved coming to school in morning and seeing all the helmeted bikers riding at a good clip together" and "Hope we do it again".

## In Conclusion

The cycling teachers and PE teachers at each school all noted that they were sorry, just like the students, when the Specialized cycling program had ended. Most participants, students and teachers alike, noted how they wanted to continue. The cycling lead teacher at Wilson summed it up best by saying to one of the scientist that “I missed my little buddies, I really miss them. We are already talking about starting up first thing in the spring. We are going to open it up and try to get 100 kids, maybe 200. We all became a family, a family of riders. WE ARE the CYCLERS. We do it together!”

### **Program Differences at the Two Schools (Wilson vs. Kennedy)**

The average age of students at Wilson was 12.04 years, and at Kennedy 12.28 years. The difference in ages was non-significant. There was also no difference in gender for the participants of the intervention at the two schools (5 girls, 20 boys at Wilson, 8 girls, 21 boys at Kennedy), the chi square test was non-significant.

The average number of biking sessions attended at Wilson was 14.04 and 18.1 at Kennedy. This difference was significant  $t(27.679) = -3.482$ ,  $p < 0.005$ ,  $df$  adjusted for unequal variance due to significant Levene’s. This result is complicated by the fact that the distribution of number of biking sessions is non-normal at Wilson, but not at Kennedy. This non-normality was caused by 5 drop outs at Wilson, but none at Kennedy.

The average heart rate during exercise was greater at Kennedy (151.5 beats per minute) then it was at Wilson (138.3 beats per minute). This difference was statistically significant,  $t(52) = -2.116$ ,  $p < 0.05$ .

### **What is the Effect of Exercise Intensity and Exercise Frequency on the Results of the Bike Program?**

Correlations were calculated between two exercise measures (average heart rate during exercise, and number of exercise sessions attended) and a number of measures prior to the biking intervention and after the biking intervention. Correlations were also calculated between the two exercise measures and difference scores (post-pre). Only the correlations involving the difference scores are listed below.

Because of the large number of non-normally distributed variables, all correlations were calculated with Spearman’s Rho, unless the following conditions were satisfied: 1. the Spearman’s Rho correlation between two variables was non-significant, 2. the Pearson’s correlation between the two variables was significant, and 3. both correlated variables were normally distributed. Partial correlations controlling for the effect of school were used because it was believed that the interventions at the two schools were potentially different (significant correlations were very different at the two schools).

### Overall Partial Correlations: Exercise Measures and Pre vs Post Test Differences

Number of sessions attended and change in: waist circumference  $r_s = -.396$ ,  $p = .002$ ; weekend screen time  $r_s = .293$ ,  $p = .017$ ; and clarity of feelings  $r_s = .280$ ,  $p = .021$

Average heart rate and change in: feeling awake in the morning  $r_s = -.319$ ,  $p = 0.02$  (two tailed); Connor's screening scores  $r_s = .238$ ,  $p = .045$ ; dispositional optimism  $r_s = -.272$ ,  $p = .049$  (two tailed); reaction time on wrong responses on Connor's CPT  $r = -.255$  (Pearson's),  $p = .049$ ; and correct responses on the CPT  $r_s = -.28$ ,  $p = .034$ ; errors on Go-No Go task  $r_s = .303$ ,  $p = .026$ ; reaction time on Go-No Go task  $r_s = .333$ ,  $p = .016$ ; reaction time on congruent responses on Flanker  $r_s = -.299$ ,  $p = .024$ ; neutral  $r_s = -.226$ ,  $p = .07$ , and incongruent  $r_s = -.220$ ,  $p = .076$  responses were also close to significance.

#### ADHD Partial Correlations

Average heart rate and change in: Go-No Go Errors (for P responses)  $r_s = .532$ ,  $p = .04$  (two-tailed).

#### Non-ADHD Partial Correlations

Number of sessions attended and change in: waist circumference  $r_s = -.519$ ,  $p = .001$ ; clarity of feelings  $r_s = .387$ ,  $p = .014$ , positive affect  $r_s = -.430$ ,  $p = .032$  (two tailed);

Average heart rate and change in: flanker reaction times for all items congruent  $r_s = -.520$ ,  $p = .008$  (two tailed); neutral  $r_s = -.467$ ,  $p = .018$  (two tailed); incongruent  $r_s = -.511$ ,  $p = .010$  (two tailed); Connor's CPT errors of commission  $r_s = .392$ ,  $p = .048$  (two tailed).

#### Behavioral ADHD Partial Correlations

Average heart rate and change in: clarity of feelings  $r_s = .346$ ,  $p = .028$ ; dispositional optimism  $r_s = -.412$ ,  $p = .022$  (two tailed).

#### Not Behavioral ADHD Partial Correlations

Number of sessions attended and change in: waist circumference  $r_s = -.610$ ,  $p = .002$ ; screen time during school  $r_s = -.449$ ,  $p = .046$  (two tailed); week-end sleep  $r_s = .434$ ,  $p = .028$ ; exercise enjoyment  $r_s = .424$ ,  $p = .031$ ; awakeness on mornings of exercise  $r_s = .532$ ,  $p = .008$ ; positive affect  $r_s = -.601$ ,  $p = .01$  (two tailed); negative affect  $r_s = -.420$ ,  $p = .046$ ; clarity of feelings  $r_s = .480$ ,  $p = .016$ ; flanker reaction times for all items congruent  $r_s = .513$ ,  $p = .021$ ; neutral  $r_s = .464$ ,  $p = .035$ ; incongruent  $r_s = .433$ ,  $p = .047$ .

Average heart rate and change in: screen time before school  $r_s = .512$ ,  $p = .022$  (two tailed); sleep on school nights  $r_s = .446$ ,  $p = .048$  (two tailed).

#### Interpretation

It is, of course, not possible to make causal inferences using the correlations presented here. Also note that a correlation's significance is given using one tailed values for the overall results simply to highlight results that might be interesting. These correlations are further examined in the group sections where a two-tailed test is used when there was no pre-testing expectation, and a one-tailed test is used for results confirming expectation. The purpose of the correlations listed here is in the generation of new hypotheses, to be tested in future empirical work, and caution should be taken in any interpretation of the results.

Some correlations were found consistently across groups. Number of sessions attended was correlated with decreases in waist circumference pre to post exercise; the probable interpretation is that biking led to these changes. Note that number of sessions attended was also negatively correlated with pre exercise obesity measures. One interpretation of this result is that more obese people liked biking less (the negative correlation between pre exercise waist circumference and exercise enjoyment was significant for ADHD subjects only), and found it harder to participate. Obesity measures are analyzed experimentally later in this report.

The correlation between number of sessions attended and week-end screen time is interesting. Screen time before school decreased significantly post vs. pre during the biking intervention, and this correlation suggests that the missed time in front of the TV in the morning is made up for on the week-end.

The correlations between average heart rate and other variables were problematic. There were two issues surrounding the variable of average heart rate. The first issue was that although heart rate measures were designed to be taken throughout the program, teachers didn't record the data after an initial period where heart rates were recorded. Secondly, heart rate zones were calculated for subjects and, to control for cardiovascular fitness levels, time in target heart rate zones was meant to be the measure used in correlations. However, only Wilson school was able to produce these measures, and so these correlations are not available.

The link between clarity of feelings and both average heart rate and number of sessions attended is very interesting. Note that the original overall Pearson correlation between average heart rate and clarity of feelings was highly significant, ( $r=.392$ ,  $p<.005$ ) and became non-significant only when controlling for school. A partial correlation was probably not necessary in this case, since the correlation between the two variables was similar at the two schools (.3 and .359). A correlation in the .3 to .4 range is considered a moderate effect, and, for example, a correlation of .392 would explain around 15% of the variance between the two variables.

People can misattribute their arousal; the classic study by Dutton and Aron (1974) had men more likely to later phone an attractive woman if they were met by her immediately after crossing a bridge than if they were met some time afterwards. The interpretation of this experiment was that the arousal that occurred in the men in crossing a high bridge was later misattributed to the attractiveness of the woman. In the current study, it is possible that exercise is teaching participants how to discriminate causes of arousal. The more intense a person exercises, the higher their arousal level, and the more training the person gets in discriminating sources of arousal that are not emotional in cause, which leads to greater clarity of feelings. It is also possible that people who are better able to discriminate between sources of arousal are better able to tolerate higher intensity exercise, and so tend to exercise harder. It is also interesting that when broader diagnostic categories were used, only ADHD subjects demonstrated a significant correlation between exercise intensity and clarity of feelings. It seems there is some link between ADHD symptoms, attribution of arousal, and exercise intensity tolerance.

It is also interesting that for non ADHD subjects, number of sessions was significantly linked to clarity of feelings. It may be that more experience with moderately intense exercise is improving their ability to understand arousal underlying feelings, but, again, a reverse causal explanation is possible. One hypothesis explaining all of these results is that ADHD subjects require intense levels of arousal to make cognitive judgments surrounding feelings, whereas normal students require only moderate intensity, and so, for them, the number of exercise periods is more important than the intensity. This hypothesis is only preliminary, but it would be interesting to test it experimentally.

### **Do Students Improve after Biking (Pre vs. Post Comparison)**

The difference between all variables pre vs. post was examined. Note that for the Kennedy school, two choices of premeasures were available, and the earlier one was always chosen (because it occurred at the same time as the Wilson pre measures). The data was analyzed in the following manner. First, a Wilcoxon Signed Rank test was performed on the pre vs. post exercise differences. Secondly a dependent t-test was performed on all the differences. If scores were significant using the t-test, but not the Wilcoxon, then tests of normality were run on the variable distributions, and if assumptions of normality were not violated, these t-tests were reported. Otherwise, all of the z scores reported below are associated with Wilcoxon Signed Rank tests. Finally, any ADHD groups reported below involve the behavioral definition of ADHD subjects (parent reported child difficulties in attention), and not the diagnostic definition (formal ADHD diagnosis).

#### Fitness Data

There were significant changes in BMI, waist circumference, and Pacer time. Average BMI's decreased from 19.8 to 19.57 ( $z=-2.3$ ,  $p<0.05$ ). Average waist circumference decreased from 28.1 inches to 27.5 inches ( $z=-3.51$ ,  $p<0.001$ ). Pacer times increased from an average of level 5 to an average of level 6, ( $z=-2.33$ ,  $p < 0.05$ ). When analysis was split between behavioral ADHD and non ADHD categories, ADHD subjects demonstrated significant changes in waist 28.9 inches pre to 28.2 inches post ( $z=-3.02$ ,  $p<.005$ ) and Pacer level 4.3 pre to 5.1 post ( $z=-2.47$   $p<.05$ ).

Balance also significantly improved after biking. Dependent t tests revealed overall balance improvements from 84.7 to 91 in the two feet eyes open condition ( $t(48)=-3$ ,  $p<0.01$ ), from 68.4 to 76.6 in the right foot eyes open condition ( $t(48)=-2.91$ ,  $p<0.01$ ), and from 66.1 to 75 in the left foot eyes open condition ( $t(48)=-2.51$ ,  $p=0.016$ ). It is assumed that balance in the eyes closed conditions did not improve since biking was done with eyes open.

#### Intervention Influences on Non-Intervention Variables

Two variables changed after the biking intervention. The first was screen time before school, which changed from an average of 12.78 hours a week to 6.67 hours a week ( $z=-2.26$ ,  $p < 0.05$ ). This change was almost certainly due to biking occurring during the time subjects would be watching TV or on a computer, and is hypothesized to have a positive effect on attention and academics if it has any effect at all. It is thus possible that a potential mechanism of exercise effectiveness is not through biking itself, but by a reduction of screen time that occurred as a result of the biking program. However, the

total hours of screen time per week was not significantly reduced as a result of the biking program, so this explanation is unlikely.

Physical activity hours outside of the biking intervention decreased pre vs. post, from 11.83 to 9.81 hours ( $t=1.937$ ,  $p=0.029$ ). Note that both physical activity variables demonstrate a rightward skew and were significantly non-normal, making this t-test questionable, and also note that the Wilcoxon for this difference was non-significant. When the hours spent biking were added to the estimates of physical activity hours outside of biking, the total hours of physical activity per day rose from 11.83 hours to 12.31, a very modest, statistically non-significant, increase. Self-report of physical activity is subject to bias, and seasonal effects may have influenced the results, but it appears likely that physical activity did not increase as much as expected during the biking intervention due to a decrease in outside physical activity.

### Affect

There were no significant pre vs. post differences in affect. For ADHD subjects only, scores decreased on the measure of attention to feelings ( $z=-1.99$ ,  $p<0.05$ ).

### Attention

Connor's attention screen scores decreased from 7.96 to 7.37 ( $z = -1.649$ ,  $p=0.049$ ), indicating fewer attention problems after biking. It is hard to come up with an explanation for these results other than the biking intervention, but alternative explanations cannot be ruled out at this point in the analysis. These score differences were significant only for non-ADHD subjects, though, 6.7 pre to 5.1 post ( $z=-2.28$ ,  $p <0.05$ ).

### PEBL Connor's CPT

Three measures changed pre vs. post on the Connor's: foil accuracy rate .334 pre to .413 post ( $z=-2.93$   $p<0.005$ ), commission errors 24.4 pre to 21 post ( $z=-3.17$ ,  $p<0.005$ ), and average correct response reaction times 403.7 ms pre to 415.9 ms post ( $z=2.68$ ,  $p < 0.01$ ). These scores overall are indicative of subjects improving their accuracy by taking more time to respond, thus decreasing the rate at which they erroneously respond to the X. This statistic does not indicate if the change is due to exercise, or greater experience when taking the test a second time. ADHD subjects demonstrated a large number of changes post intervention, in target accuracy, commission errors from 24.7 to 21.3 ( $z=-2.28$   $p<0.05$ ), omission errors 25.8 to 41.7, 14.5 pre to 26.1 post after removing outliers, ( $z=-2.093$ ,  $p <0.05$ ), (foil accuracy  $z=-1.628$   $p=0.052$  one tailed just missed significance) and correct reaction time mean 396 ms pre to 428 ms post ( $z=-2.248$   $p<0.05$ ). ADHD subjects appeared to be taking longer in order to improve commission errors rates, while also making more errors of omission in doing so.

### Go-No Go

Response times also increased on the Go No Go test ( $z=-2.252$ ,  $p< 0.05$ ). The decrease in total errors on the test just missed significance ( $z=-1.589$ ,  $p=0.066$ , one-tailed). ADHD subjects took longer in

responding the second time around 560ms pre to 652 ms post ( $z=-3.67$   $p < 0.001$ ), but did not demonstrate significant accuracy gains in doing so.

### Flanker

No flanker scores were significantly different pre vs. post, but all scores were in the direction of an increase in reaction times and an increase in accuracy. For ADHD subjects only, Flanker congruent accuracy scores improved pre .76 to post .86 ( $z = -1.828$ ,  $p < 0.05$ )

## **Can We Attribute Changes To The Biking Program?**

### **Long Term Changes as a Result of Biking**

Results after the month long biking program are given below, with interesting results graphed. Appendix B contains a larger number of graphs that illustrate the results of testing more comprehensively.

All long term results for a particular variable were run in a mixed model ANOVA where scores of that variable at the three measurement times were the within subjects variable, and school and ADHD diagnosis were the between subjects variables. Normality of variables was tested using Kolmogorov-Smirnov, and Levene's and Mauchly's tests were run as well to test variance assumptions. These tests can be assumed to be non-significant in the results presentation below unless otherwise stated. In instances where non-parametric tests are used, normality can be assumed to have been violated according to Kolmogorov-Smirnov.

In interpreting graphs of mixed model results, it should be remembered that Wilson school did the biking intervention from September to October, and Kennedy from October to November. Thus, hypothesized improvements should occur for these time periods for the two schools. The September to October time period for Kennedy can be examined to give an indication of what the effect of going back to school might be on student scores on a particular measure, and the October to November time period for Wilson can be examined to determine how long-lasting changes due to biking were.

### BMI and Waist Circumference

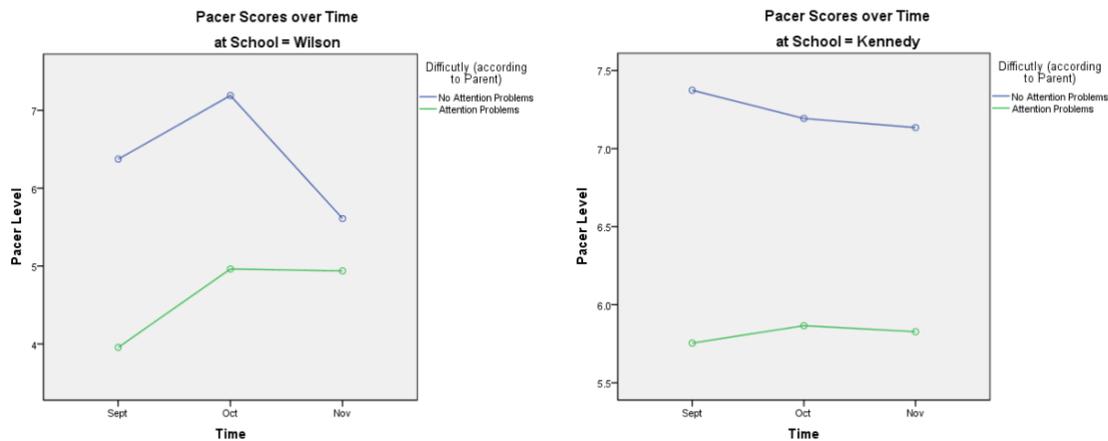
All subjects at both schools demonstrated a decrease in waist circumference after biking; there was a main effect of time  $F(2,98)=8.93$ ,  $p < .001$ . All statistics were non-significant for the BMI analysis, including the main effect of time. The only trend that stood out in the data was that the BMI's of non-ADHD students at Wilson school appeared to increase, whereas BMI's appeared to decrease in every other condition. This trend is interesting because waist circumferences appear to be decreasing for this same group during biking but not after biking. A non-increase in waist circumference along with an increase in BMI probably indicates muscle gains made during the time period.

The lack of a main effect of BMI is not problematic, since the analysis includes time periods before and after the biking intervention that are of less interest for cardiovascular fitness variables (control groups are less needed since cardiovascular improvements after exercise are a well-established

finding, and causal claims are unlikely to be disputed). The pre vs. post difference did not include these less relevant time periods, and was significant. A more important question is how useful the BMI and waist circumference changes are. A decrease in BMI from 19.8 to 19.57 pre to post may be statistically significant, but it seems unlikely that it represents a change that will lead to any long term benefits to a child. A biking program that was of longer duration and that had follow-up measurement points at longer intervals would be needed to demonstrate any stable benefits of the biking program in reducing obesity.

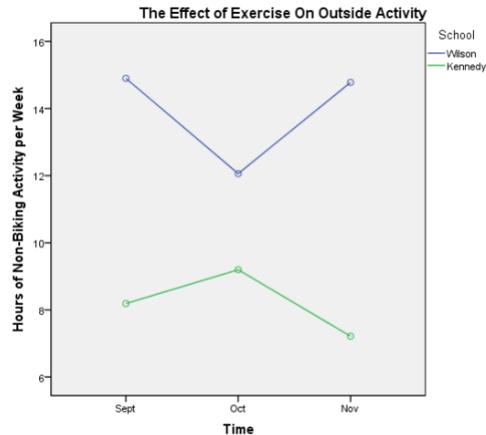
### Pacer Scores

There was a significant overall difference pre vs. post in pacer scores, indicating that on average, subject's cardiovascular health was improving. When the mixed model analysis was run, significant differences were found between schools, and between ADHD and non-ADHD subjects. There was a significant time by diagnosis interaction ( $F(2,78)=4.221, p<0.05$ ). Simple main effects revealed that for ADHD subjects, September Pacer scores were lower than October Pacer scores, and the difference between September and November Pacer scores just missed significance. There were no differences over time for non-ADHD subjects in Pacer scores. There was also a significant time by school interaction ( $F(2,78)=4.681, p<0.05$ ). Simple main effects were run and Pacer scores were shown to be significantly different between September and October at Wilson school only.



### Hours of Physical Activity Outside of Biking

Pre vs. Post measures revealed a significant difference between outside physical activity before the intervention compared to after the intervention: subjects did less outside physical activity while biking. There were a number of issues with this result, including a few extreme outliers (one subject reported doing 43 hours of exercise outside of biking, a highly unlikely number) and non-normal distribution of scores. However, a more detailed analysis by schools reveals that biking almost certainly led to less outside exercise. A time by school mixed model analysis demonstrated a significant time by school interaction, and specifically, a significant time by school quadratic interaction  $F(1,51) = 6.612, p<0.05$ . Looking at the graph below, it is apparent that exercise decreased during the intervention at Wilson in October, but returned to normal levels in November when biking had stopped. Activity levels showed a similar kind of decrease at Kennedy when biking occurred in November.



These results indicate that a different kind of intervention may have occurred than was planned. The biking intervention was designed to occur alongside student's normal physical activities. The study was thus meant to be an examination of an **additional** biking activity on top of normal exercise. What the study ended up being according to student self-report was an examination of a program that kept the total amount of exercise the same, but changed the mode, intensity, and time of exercise for two and a half hours a week – the biking intervention.

### Exercise Enjoyment

ADHD students liked exercise less than non-ADHD students. For example, using behavioral measures of ADHD, and self-reported exercise enjoyment ratings from the questionnaire, an overall effect was found for ADHD subjects across all time periods  $F(1,49) = 4.911, p < 0.05$  (ADHD subjects liked exercise less). Because this result could have been potentially influenced by non-normality of the data, group differences at specific time points were examined using Mann-Whitney U tests. ADHD students were found to like exercise less than non-ADHD students in October ( $U=235, z=-2.09, p < 0.05$ ) and November ( $U=244, z=-1.92, p < 0.05$ ) (both tests one-tailed).

### Getting Along with Classmates

There were no changes in how much students liked each other due to the biking intervention. Student's ratings of classmate liking were in general high, so there may have been a ceiling effect. It is noteworthy that ADHD students (behavioral) liked their peers less than non-ADHD students in September (Mann-Whitney  $U=242, z=-1.92, p < 0.05$ ), and November (Mann-Whitney  $U=214.5, z=-2.55, p < 0.01$ , both tests one-tailed).

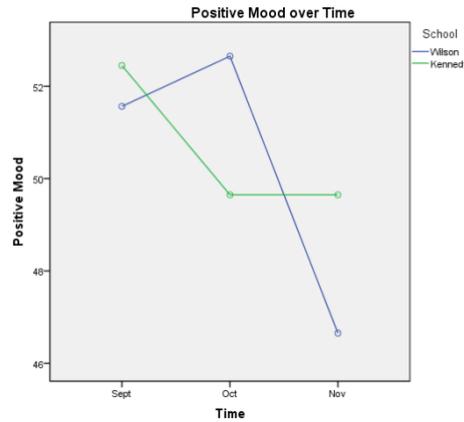
### Sleep and Awakeness

One potential issue with the study was that students would get less sleep due to waking up early for their morning biking intervention. Fortunately, there were no significant changes in self-reported awakeness in the morning according to a large number of different statistical analyses. There were also no changes in self-reported average sleep on week-days either, and no students in any group in any analysis ever reported getting less than 8 hours of sleep a night at any time point (except that ADHD

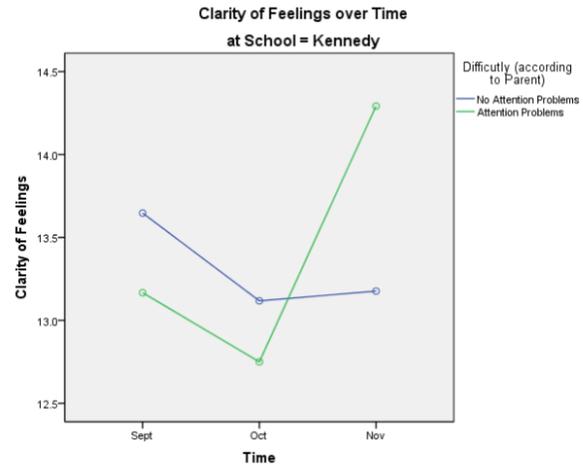
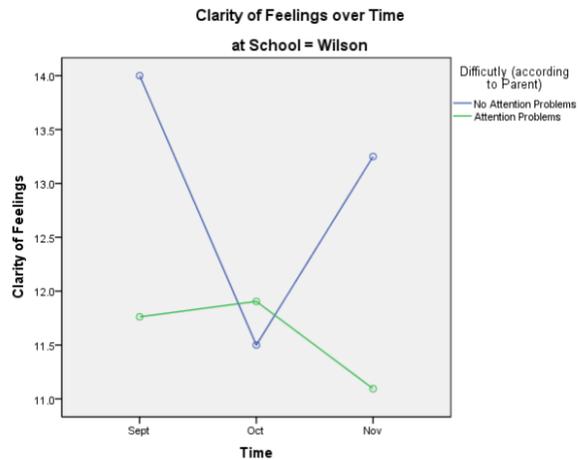
students at Kennedy school reported an average of 7.9 hours of sleep before the biking intervention started in September). There was also a significant difference between sleep on the week-end and sleep on week-days (Sept 8.6 weekday vs. 10.3 week end hours, Oct 8.6 vs. 10.3, Nov 8.5 vs. 10.2). These values were all significant below  $p < 0.001$  for both parametric and non-parametric tests. This difference, combined with a median awakeness rating of 2 on a scale of 1 to 5 for all groups and time periods, suggests that students may not have been getting enough sleep on school days, before, during, and after the biking intervention.

### Affect

There was a significant school by time interaction in the quadratic trend for positive mood scores on the child PANAS ( $F(1,40)=7.72, p<0.01$ ). This interaction can be explained entirely in terms of the biking intervention, as scores tended to increase or stay the same during the intervention, and decrease when the intervention was not being run. Biking appears to improve mood, or at least prevent its decrease.



There was a significant school by behavioral diagnosis by time interaction quadratic trend for clarity of feelings ( $F(1,50)= 4.1, p<0.05$ ). This is an unexpected trend, as it appears that three of the groups demonstrated increased clarity of feelings during the biking intervention in comparison to after the interventions, but the non-ADHD group at Wilson appears to show much lower clarity of feeling at the end of the intervention time period in comparison to before or afterwards. It appears that biking may lead to greater clarity of feelings (at least self-reported). However it is difficult to come up with a reason why the non-ADHD Wilson students decreased so much in the self-reported clarity with which they were able to understand their own emotional responses at the end of the biking intervention.

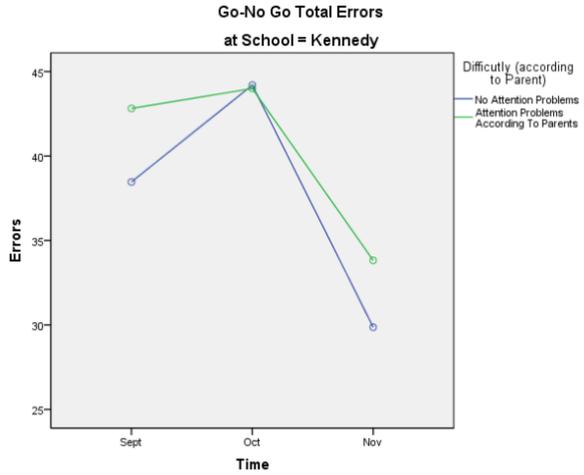
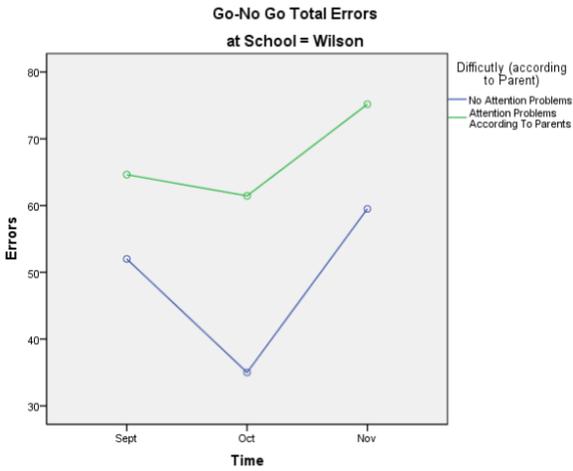


There was a significant time by school linear trend for mood repair ( $F(1,50)=4.1, p<0.05$ ). Student's self-reported ability to repair their mood increased at Kennedy school over time, and decreased at Wilson. There is no evidence that biking was responsible for this result.

There were no statistically significant results for dispositional optimism scores, or for attention to feelings.

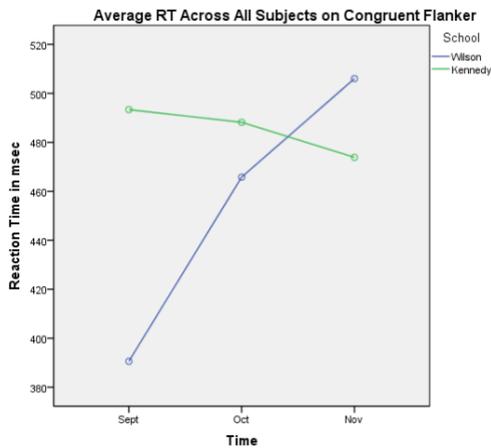
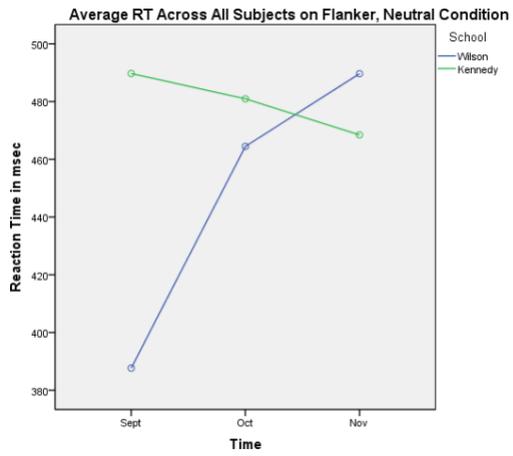
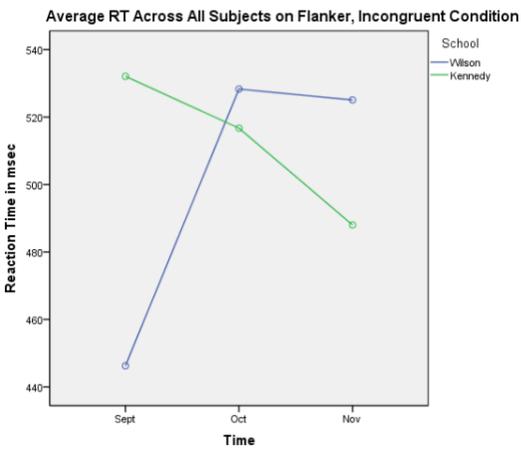
### Go No Go Tests

There was a significant quadratic trend in the time x school interaction for total errors on the Go No Go task  $F(1,35) = 4.98, p < 0.05$ . Students committed fewer errors at the end of the biking intervention in comparison to measurement times before the intervention, or a month afterwards. It appears as if biking improves student's test accuracy, a function of attention and response suppression. Student's average reaction times generally increased on the test, there was a significant linear trend  $F(1,35)=6.79, p<0.05$ . This overall linear does not seem to capture the performance of specific groups within schools and diagnostic categories, though, and caution should be used in interpreting this result broadly. Note that subjects also performed the Go No Go test in the middle of the biking program for acute testing. This result is not added in to the long term testing results because half the subjects took the test after biking during acute testing, and none of the subjects did this for any of the long term testing measurements.



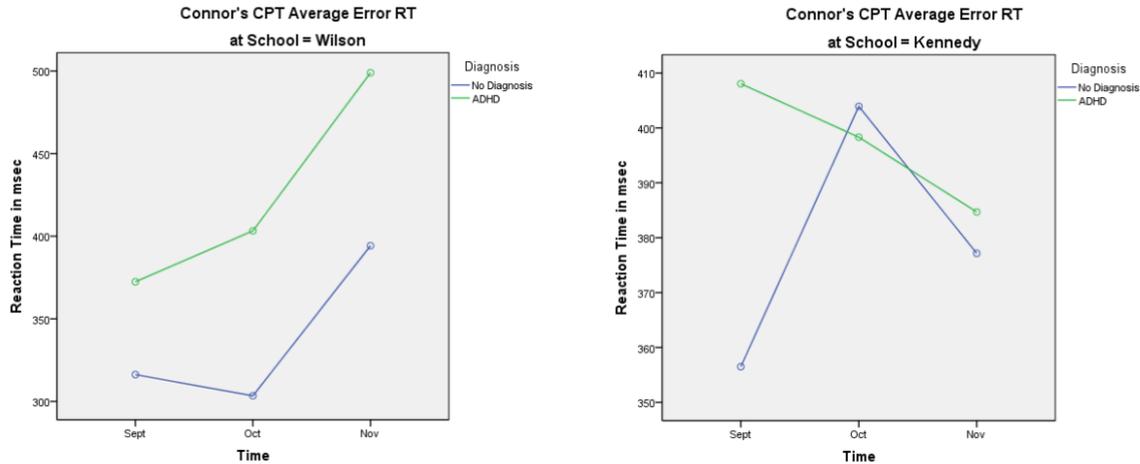
Flanker

There was a significant time x school interaction on incongruent trials for RT,  $F(2,72) = 5.191$ ,  $p < 0.01$ . This interaction also existed for neutral trials ( $F(2,72) = 5.721$ ,  $p = 0.005$ ), and for congruent trials ( $F(2,70) = 3.474$ ,  $p < 0.05$ ).



## PEBL Connor's CPT

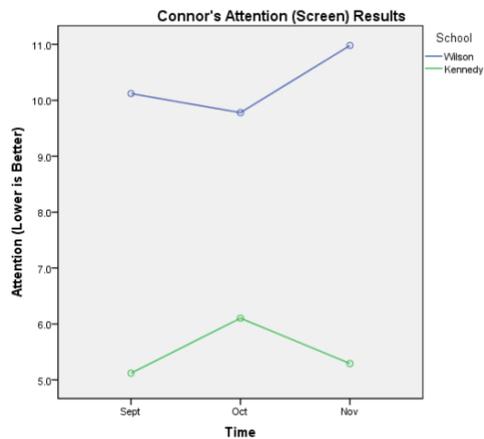
On the Connor's CPT, there was a significant time x school linear trend for reaction time on the incorrect responses ( $F(1,36)=7.53, p<0.01$ ), and a significant time x school quadratic trend on the correct responses ( $F(1,36)=5.2, p<0.05$ ).



There were no significant effects on response accuracy. Examining the graphs, the trend appears to be that accuracy decreased over time on the Connor's at Wilson, and decreased at Kennedy before exercise and increased/stayed the same afterwards.

## Connor's Screening Test

Although the Connor's Screening test scores were significant pre vs. post, the quadratic trend just missed significance. The trends appear to be in the right direction, though.



## **Acute Changes Due to Biking (Changes that Occur Immediately After Biking, on top of the Long Term Changes)**

### Symbol Digit, Five Point Test, Animal Naming, & FAS Test

Test results were significantly correlated with one another amongst all four tests. Testing after biking did not have a significant effect on test scores, both for overall groups, and for subgroups comprised of ADHD (defined by either diagnosis or parent report) and normal students. (There were significant effects of diagnosis, however, as ADHD subjects always performed significantly worse). However, there was an interesting trend. For ADHD subjects, on the symbol digit and five point test, subjects scores were higher in the biking condition than in the non-biking condition. On the animal naming and FAS tests, the scores were lower in the biking condition than in the non-biking condition. There are two explanations for this potential effect. The first is that there is no actual difference between the biking and non-biking scores. The second is that there is a difference that did not reach statistical significance. The tasks differ in an interesting way. The animal naming and FAS tests require subjects to generate responses and then to detect whether or not the response had been given before. Performance can decrease due not only to generating fewer responses in a given time period, but by generating too many responses that have already been given. The symbol digit and five point test can be conducted systematically so that interference from prior responses is much less of an issue. It would be interesting to determine if ADHD subjects tend to generate previously given responses more than non-ADHD subjects, or to examine whether they have more difficulty resolving this kind of conflict.

Three trends approached significance – there was almost an overall level of significance across all tests in the MANOVA for biking (Pillai's trace(6,35)=2.173,  $p=0.069$ ). The ANOVA for biking for the five point test  $F(1,40)=3.65$ ,  $p=0.063$ ) was also close as was a biking x diagnosis interaction for the symbol digit  $F(1,40)=3.26$   $p=0.078$ ).

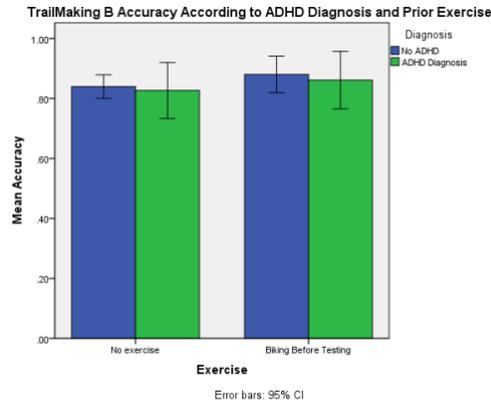
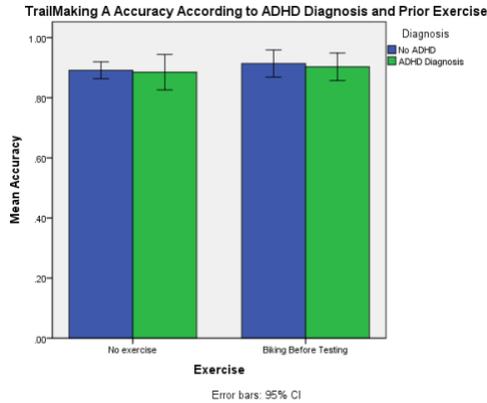
#### Go-No Go Test

All scores were non-significant in biking and non-biking conditions for all groups of subjects, and for both parametric and non-parametric tests. The trends were in the expected direction, bikers performed the task faster and made fewer errors, both overall and for ADHD groups. However, none of these differences reached statistical significance.

#### PEBL TrailMaking Test

Both average number of clicks and average accuracy measures turned out to be non-normally distributed (as indicated by both Kolmogorov-Smirnov tests and Shapiro-Wilk tests), and looking at histograms and Q-Q plots, results were hugely negatively skewed for accuracy and hugely positively skewed for average number of clicks, especially for biking before test groups and ADHD groups. Parametric testing was therefore not done, and the z scores below represent Mann-Whitney U testing.

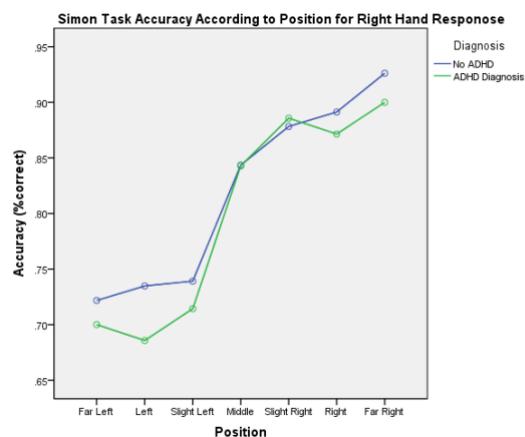
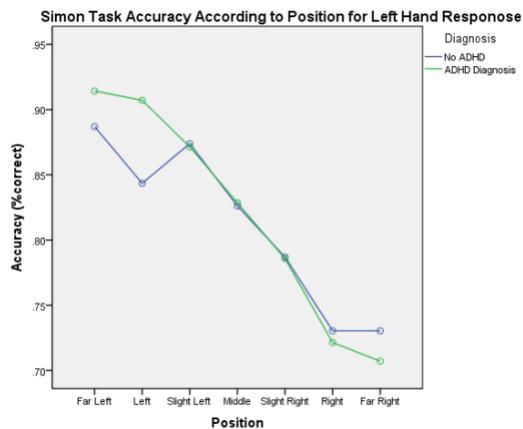
To examine the interaction between biking vs. no biking before test, and diagnosis vs. no diagnosis, ANOVA's were run for Trail Making A and B accuracy and median response times per item. Biking improved Trail Making accuracy overall for both groups on the Trail Making A ( $z = -2.257$ ,  $p < 0.024$ ), and Trail Making B ( $z = -2.08$ ,  $p = 0.045$ ). There was no significant effect of biking on reaction times.



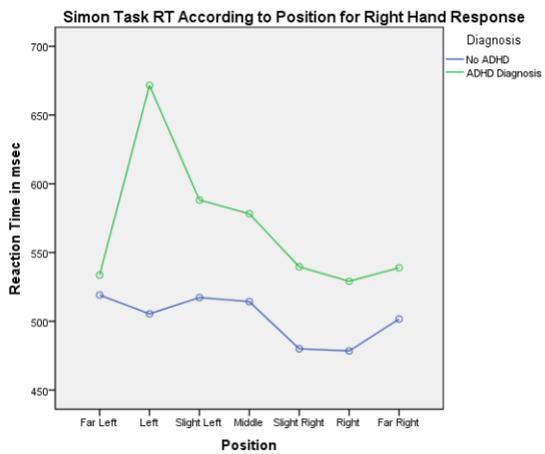
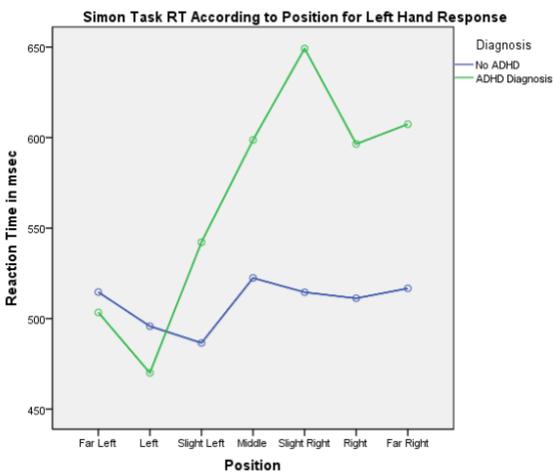
### Simon Test

The Simon test is often analyzed using a difference score. The mean RT or percent correct on congruent trials are subtracted from the RT or percent correct on incongruent trials. In the PEBL Simon used in the current study, circles are presented a certain amount to the left or right of center (200, 100, 50, or 0 units, with the exact distances determined in part by the size of the computer display used). Instead of two different congruence conditions (congruent vs. incongruent), seven different conditions thus exist (labeled in the graphs according to distance from center – far left, left, slightly left, middle, slightly right, right, far right). Analysis was performed on all seven conditions using a mixed model design with presentation location as a within subjects variable, and ADHD diagnosis (yes or no) and biking exercise before testing (yes or no) as between subjects variables. Polynomial trends were also calculated.

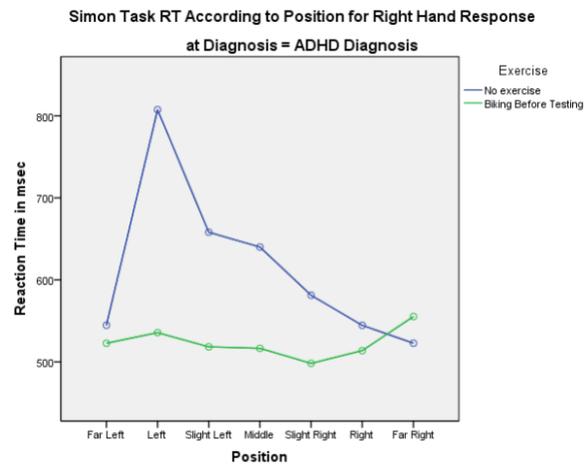
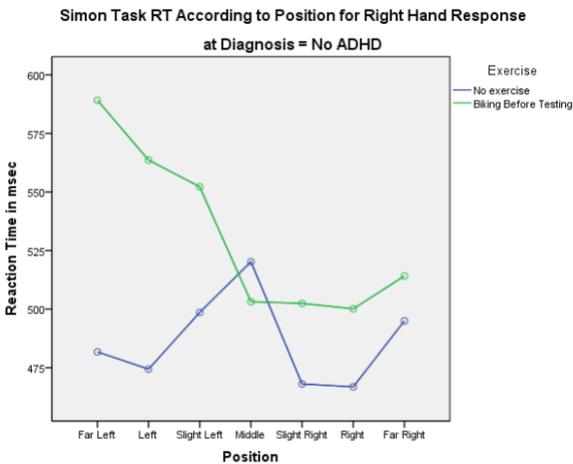
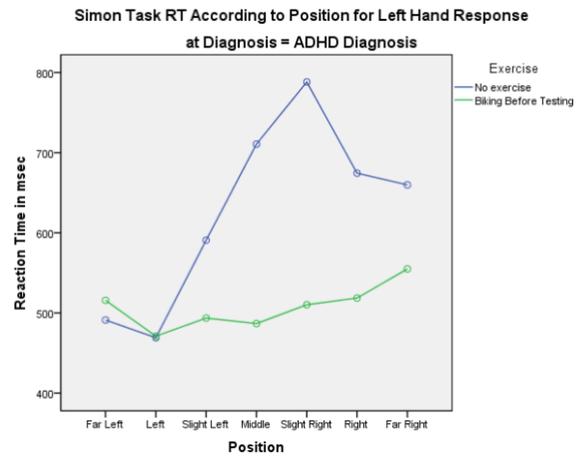
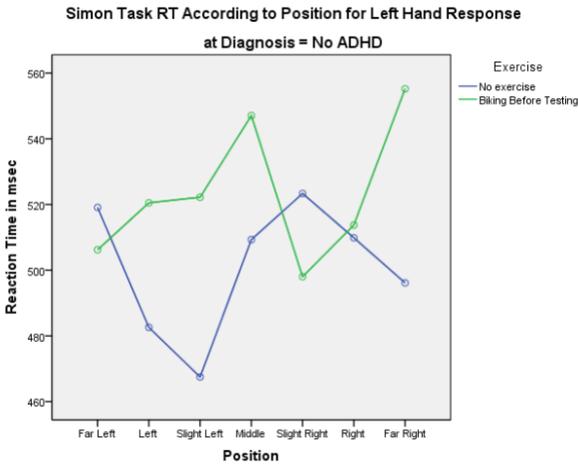
These results and the accompanying graphs can be hard to interpret, so first the effect of position will be considered to demonstrate the general effect on the Simon task. For response accuracy there were significant main effect of position: left hand responses  $F(6, 210) = 14.308, p < 0.001$ ; right hand responses  $F(4.473, 156.554) = 17.573, p < 0.001$ , df Greenhouse-Geisser corrected due to Mauchly's test of Sphericity  $< .75, p < 0.05$ . The linear trend for response accuracy was also significant: left hand responses  $F(1, 35) = 62.593, p < 0.001$ ; right hand  $F(1, 35) = 53.408, p < 0.001$ . If the graphs accompanying the stats are examined, then it can clearly be seen that as presentation/response sides become more incongruent, accuracy decreases in a linear fashion, for both ADHD and non-ADHD subjects.



Similar results occur for reaction time, as incongruence increases, RTs increase linearly as well. Main effect of position: left hand response  $F(3.152, 110.319) = 2.997$ ,  $p < 0.05$ , df Greenhouse-Geisser corrected due to Mauchly's test of Sphericity  $< .75$ ,  $p < 0.05$ ; right hand response ns. There is a linear trend for both left side responses  $F(1, 35) = 12.169$ ,  $p = 0.001$ , and right side responses  $F(1, 35) = 5.070$ ,  $p < 0.05$ . Looking at the accompanying graphs, however, shows that perhaps things are not so simple. It looks like ADHD subjects may perform differently than non-ADHD subjects. For right hand responses, this position x diagnosis hypothesis is not backed up by inferential statistics, but for left hand responses, the position x diagnosis interaction approaches significance, and more importantly, there is a position x diagnosis linear trend. Looking at the graph for right hand responses, we can see that the ADHD and non-ADHD lines look different here as well, but for some strange reason ADHD students did relatively well at the most incongruent condition. A working hypothesis based on these results is that accuracy decreases as congruency decreases on the Simon test, but that RTs increase as congruency decreases for ADHD subjects, and to a greater extent than for non-ADHD subjects.

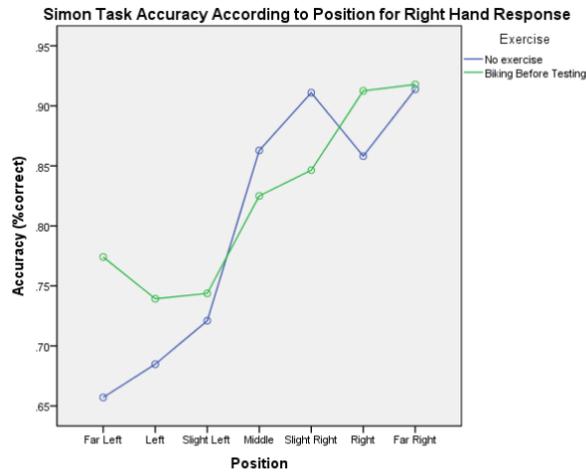
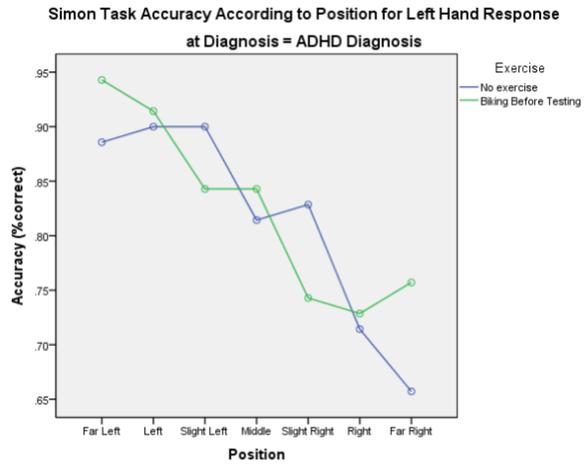
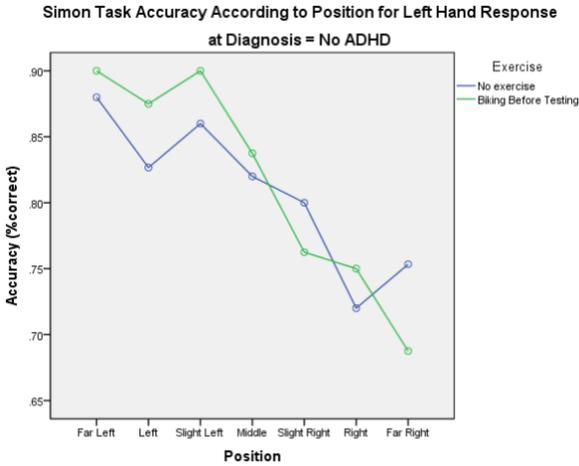


The full statistical analysis found a number of significant results. There was a significant position by diagnosis by biking linear trend for reaction times for both left hand responses  $F(1, 33) = 5.789$ ,  $p < 0.05$ , and right hand responses  $F(1, 33) = 8.146$ ,  $p = 0.007$ . Examining the graphs of reaction time results, it appears that non-ADHD subjects and post-exercise ADHD subjects perform similarly, but that ADHD subjects do significantly worse if they have not exercised before performing the Simon task.



It may appear that there is a lot of variance in the non-ADHD graphs, but if the scale of the graph is examined (the vertical axis), it can be seen that the lines would be relatively straight if plotted on the corresponding ADHD graphs (the ADHD graphs Testing cover a much larger range, from 500 ms to 800 ms).

For accuracy, there was a position by diagnosis by exercise quadratic trend for left hand responses that just missed statistical significance  $F(1,33) = 4.083, p = 0.051$ , and a significant position by exercise quadratic trend that was significant for right hand responses  $F(1,33) = 6.165, p = 0.018$ . This quadratic trend is hard to interpret, it means generally that the data for subjects can be modeled using lines with one bend in them, and that the lines are different from each other for the different groups, or that some groups can be modeled using lines with one bend in them, and some groups can't. Examining only the ADHD subjects, it looks as if exercise improves accuracy in comparison to task performance without exercise in incongruent conditions, and it looks like the bend in the line may occur when items start being presented towards the center of the screen.



## ERP Results

The results presented below are ERP results from the group of students who participated in EEG measures and completed the biking program. QEEG results are given in Appendix C, and ERP results for two different subgroups are given in Appendix D.

## Go-No Go Task Behavioral Performance

The behavioral components of the Go-NoGo picture task are described below for all bikers (N=19) at the pre measure, along with the post measure and database comparison information. Pre and post measures were compared to the HBI normative database using a t-test, and significant results indicate a difference between the scores of the ERP biking group and a group of similar children who represent typical behavior on that measure.

Bikers (N=19)	Pre Measure	Level of Significance at Pre Measure	Post Measure	Level of Significance at Post Measure	Norms for 12-13 year olds
Reaction Time (Variance)	376ms (19.5ms)	N.S. (variance N.S.)	330ms (12.4ms)	P<0.01 <b>(FASTER)</b> (variance N.S.)	457ms (11.7ms)

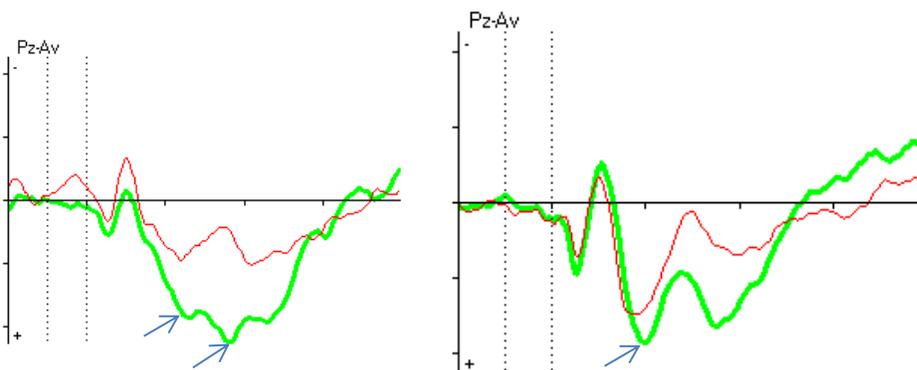
Errors of Commission for NoGo/animal-plant pairs	6.1%	N.S.	3.3%	N.S.	1.4%
Errors of Commission for plant-plant pairs	3.5%	N.S.	1.6%	N.S.	0.3%
Errors of Commission for plant-human pairs	3%	N.S.	2%	N.S.	0.2%

N.S.= Not Significant

### Go-NoGo Task Event Related Potentials:

#### Stimulus 1: P3b at Pz Measure for Bikers

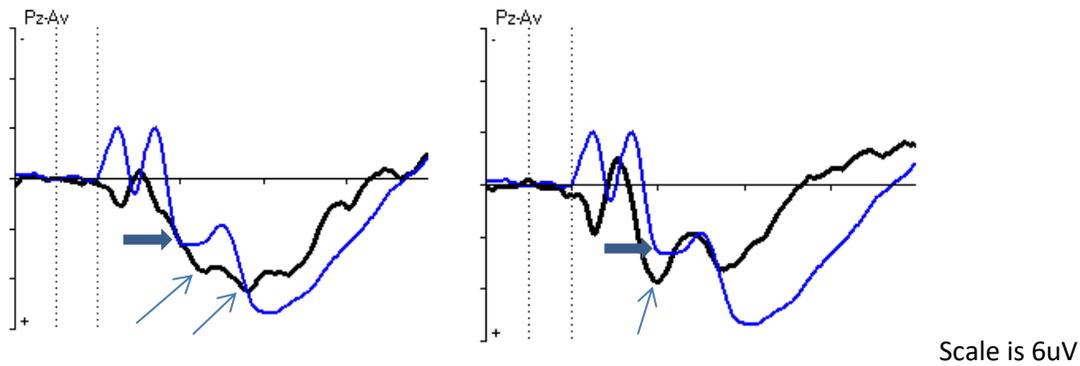
The neuroelectric indices for regular bikers are analyzed below. In the figure below, the “go” averaged trials, meaning trials when the second stimulus is an animal and subjects should push the “go” button, is shown in green, and the “no go” averaged trials, meaning the presentation of a plant as the second stimulus, is shown in red. The P3b ERP for Go trials appears with two peaks. It should be noted that in the presentation of all ERP figures, positives (P3s) are shown with a downward peak and negatives (N2s) are shown with an upward peak. The two blue arrows below point to the positive, P3b responses. These P3 waves occur at 352ms at 3.7uv, and 460ms at 4.5uv. The P3b response for the group of regular bikers is poorly defined. The attention cuing response for the group is slower by 44 to 152ms, and lesser amplitude by 1 to 1.8uv than the age matched normative group, not surprising given the fact that many subjects had ADHD or attention problems. At the post measure, the P3b has decreased latency, meaning the students “realized” or devoted attention resources to the second animal faster. The P3b appears well defined at 296ms, 3.7uv in the post measure whereas in the pre there were two peaks, 352ms and 460ms and these peaks were poorly defined. The amplitude of the P3b at the post measure is 3.7uv, while in the pre it was 3.7 and 4.5uv for the two peaks. The post measure is closer to the database norm P3b which is 308ms, 2.7uv. The pre measure for the regular bikers P3b in response to the first animal is shown in the graph to the left, and post measure in the graph to the right.



Scale is 4.5uV

A comparison with the normative database is shown below for the pre measure on the left and the post measure on the right for the P3b attention cuing response for Go tasks only. As noted above, there is trend towards normalization at the post measure. Below, the regular bikers average for to Go

stimulus one is shown in black and the database average is shown in blue. The thick arrow shows the P3b component for the database norm.

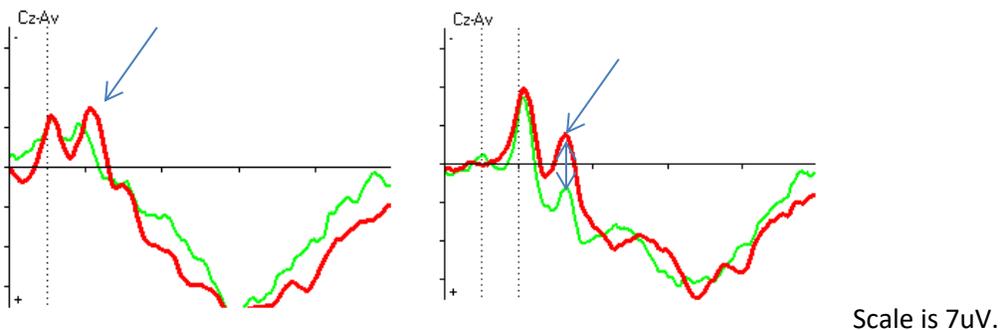


N=19	Pre Measure	Level of Significance	Post Measure	Level of Significance	Norm for 12-13 year old
P3b (Go Trials)	352ms, 3.7uV 460ms, 4.5uV	N.S.	296ms, 3.7uV	N.S.	308ms, 2.7uV

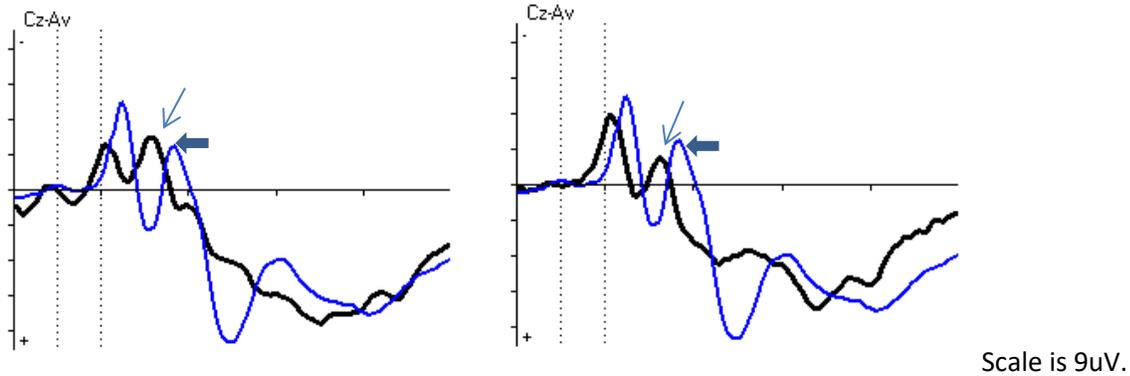
While there was no significant difference in P3b response for stimulus one in Go trials at the pre or post measure, there is a change in the expected direction, where at the post measure the P3b is close to the database norm.

**Stimulus 2: N2NOGO at Cz Measure for Bikers**

The N2NOGO component is examined as the response in the brain when the second stimulus appears and it is not an animal, after the first stimulus was an animal. The green line shows averaged Go responses, and red shows averaged No Go response. The N2NOGO for the pre measure for regular bikers is shown below to the left with a blue arrow pointing to the peak of the wave, shown in the upward direction, and the post measure is shown to the right with a similar format. Note in the post measure, there is an additional arrow indicating the difference in response between Go and No Go. This difference was not present at the pre measure. At the post measure, the inhibitory response is enhanced, a sign of the neuroelectric foundation of greater inhibitory control, despite the behavioral measures not showing the expected decrease in commission errors.



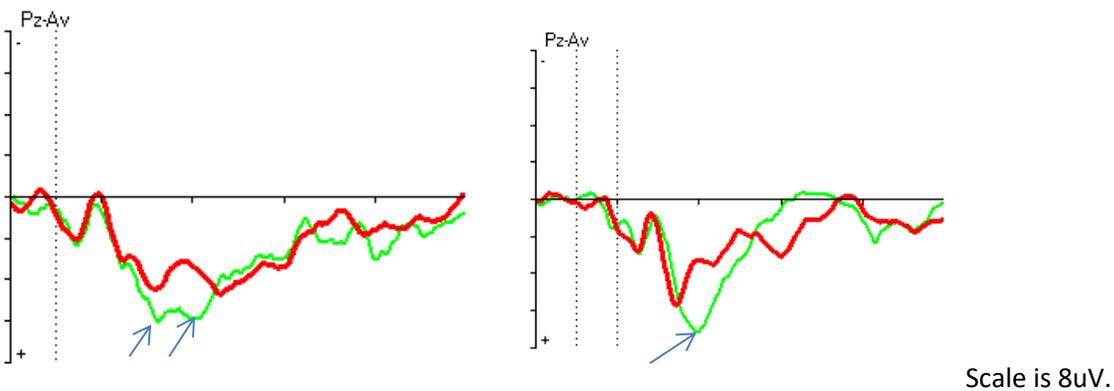
The database comparison below shows the difference between the regular bikers N2NOGO and the norm at pre (the left graph) and post (the right graph). The thick blue arrow shows the normative N2NOGO response.



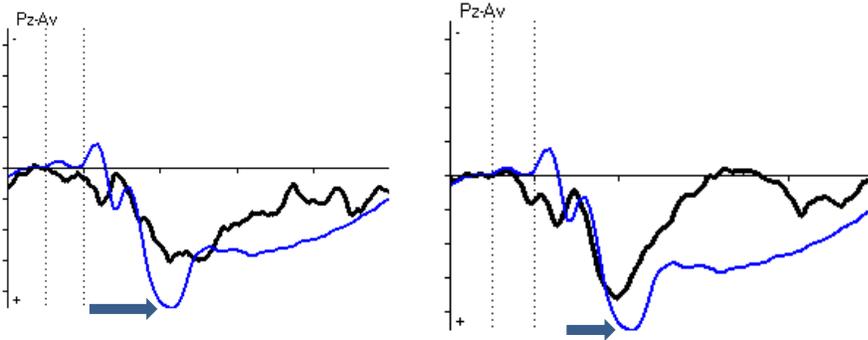
N=19	Pre Measure	Level of Significance	Post Measure	Level of Significance	Norm for 12-13 year old
N2NOGO	216ms, -3uV	N.S.	220ms, -1.5uV	N.S.	272ms, -2.4uv

### Stimulus 2: P3b at Pz Measure for Bikers

The P3b response for Go tasks is shown below at pre and post measure. The green line shows the bikers averaged Go responses, and the red line shows the bikers averaged No Go response. Note that the P3b response has two poorly defined peaks at the pre measure, and is better defined with only one peak at the post measure.



The database comparison for P3b pre and post are shown for the bikers in black and the Go task for the normative group in blue. There is no significant difference between the pre and post measures and the database, however the biker's P3b Go trend closer to the database response. The thick blue arrow shows the database norm P3b Go response below:

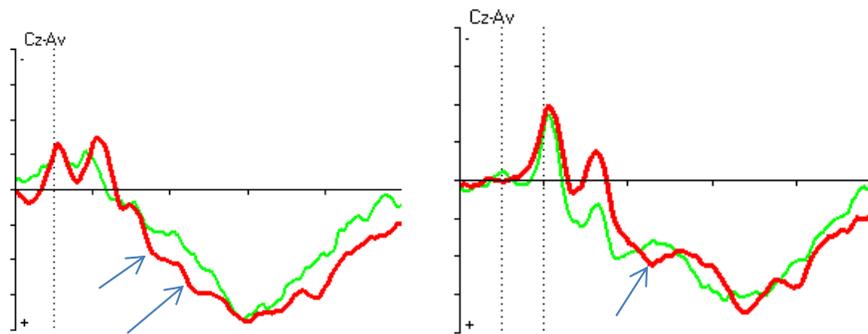


Scale is 9uV.

N=19	Pre Measure	Level of Significance	Post Measure	Level of Significance	Norm for 12-13 year old
P3b at Pz GO	324ms, 6uV 412ms, 5.9uV	N.S.	300ms, 7.1uV	N.S.	324ms, 9.1uV

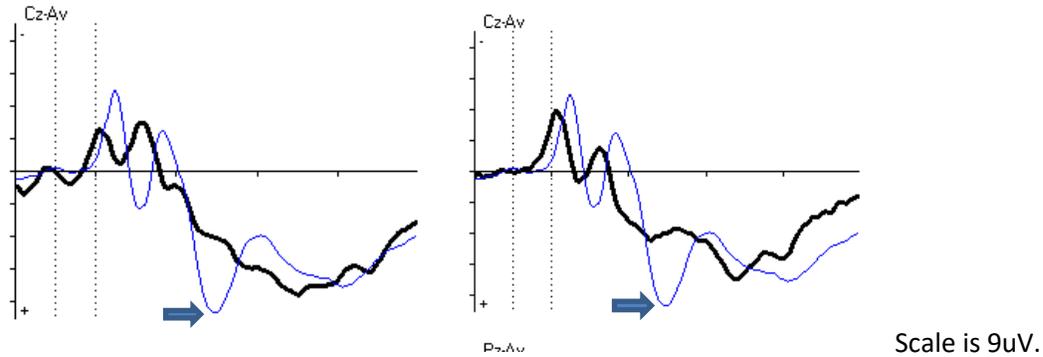
### Stimulus 2: P3NOGO at Cz Measure for Bikers

The final component examined for stimulus two is the P3NOGO at Cz. The pre (left) and post (right) measures are shown below. The P3NOGO response improves from pre to post, morphing the two averaged peaks in the pre to one well defined peak in the post.



Scale is 8uV.

The database comparison for the P3b NOGO averaged response for the bikers at pre and post is shown below with the biker's average in black and the norm in blue. While the P3b for the bikers is not significantly different from the norm, there are observable differences in the morphology of the component. At the pre measure, the signature of the inhibitory response is poorly defined, and has lesser amplitude than is usual, similar to Pontifex's 2013 study. After a maximum of 20 sessions of biking, there is a trend toward the norm: there is some definition of the P3 NOGO, but it is slightly earlier and of less amplitude than the norm.



N=19	Pre Measure	Level of Significance	Post Measure	Level of Significance	Norm for 12-13 year old
P3b NOGO at Cz	356ms, 3.9uV, 464ms, 5.7uV	N.S.	356ms, 4.5uV	N.S.	396ms, 8.6uV

## Summary

### Behavioural Results

The biking group performs the Go/No Go test quicker than the normative sample, but makes more mistakes. This is common in ADHD populations where the lack of inhibitory control leads to impulsivity.

### P3b Wave at Pz for Cue (Stimulus 1)

While the deviation at the pre measure was not statistically significant from the normative sample for the P3b wave, the movement in decreased latency and increased amplitude is in the expected direction as other publications comparing less to more cardiovascularly fit children. In the present study, increased fitness was an outcome of regular biking, and as demonstrated by a number of results from Charles Hillman's lab, fitter children have a faster, and higher P3b compared to less fit children. In the one published study (Pontifex et al., 2012) comparing the effect of a single bout of exercise on ADHD vs normal showed an increase in amplitude and decrease in latency for the P3 component following exercise. This study shows a similar effect, however the effects are not due to a recent bout of exercise as they were in the Pontifex study, rather these effects are assumed to be cumulative and associated with riding 5 days a week, for 4 weeks. It is noted that at the post measure, several days had elapsed since the final ride, ranging from 1 to 5 days following the last organized ride. Students did not ride or attend PE or recess on the days they did QEEG/ERP assessments to prevent the acute effects of exercise polluting the longitudinal changes. It is assumed that if students had biked on the days of the post neuroelectric measures, the changes would have been greater; however this is a topic for future research. An interesting question would be to address an upper fitness threshold, where a certain VO2 Max predicted smaller changes in acute information processing changes immediately after exercising, and vice versa, a minimal VO2 Max predicted the greatest changes.

### N2 Wave at Cz for No Go Responses (Stimulus 2)

The N2 response for the bikers is faster than the norm, and this may be due to inhibitory control or a potential gating response, where information processing the brain the not 'gated' correctly, and too much information is sent for processing too soon, and the networks have difficulty in the processing. The impact of the lesser amplitude in the N2/NOGO signifies fewer resources devoted to inhibitory control, however the change is not significant, and may/may not be due to chance.

#### P3 Wave at Pz for Go Responses (Stimulus 2)

Post measures were again more similar to the norm than pre measures, although statistical comparisons were non-significant. Future research should examine if a longer biking intervention (3-6 months) would continue the trend of the biker's response moving closer to the norm.

#### P3 Wave at Pz for No Go Responses (Stimulus 2)

In the pre biking measure, there is not a large difference between averaged Go and No Go responses, suggesting that the brain is showing different patterns of activation when the task requires action execution compared to action suppression. In Pontifex's study, ADHD children were found to have smaller P3 amplitude compared to the healthy match control during tasks that required greater inhibitory control. No Go tasks are more challenging to the brain than Go tasks, especially for younger age groups as the pre frontal cortex housing inhibitory control networks are not fully developed, and even more so in students with attention difficulties (whether they are identified through an ADHD diagnosis, reported parental difficulty attending, a high school on an ADHD screen). It is encouraging that these effects have been observed acutely in Pontifex's study, and over longer periods in the present study, and provide more preliminary evidence that exercise provides a benefit to students with difficulties by changing information processing at the neural level during inhibitory control tasks. A future study should determine how long it takes the P3b NOGO to 'normalize' and for inhibitory behavior to improve, also how exercise intensity, fitness status, pre measure P3b NOGO morphology and other factors co vary with improvements.

## F. Conclusions and Future Directions

### **Overview**

The one month biking program appeared to improve fitness, decrease obesity levels, lead to an increase in positive mood, lead to improvements in some aspects of attention, and lead to acute benefits in cognition on top of the long term benefits provided at the end of the program. Even when students did not score "better" on a particular test after biking, they appeared to perform differently on that test. For example, students took longer on many aspects of the cognitive tests that were performed after biking in comparison to before biking. So there are some cognitive and affective improvements that occur after biking, and we can link these improvements with biking, since they did not occur during time periods of no biking.

The improvement in the Connor's screening test results from pre vs. post represent attention improvements that are general and behavioral. Students felt that they were better able to control their impulses and keep focused on the task at hand after biking in comparison to before they participated in

the biking program. This result is of obvious benefit to teachers in schools: a student must first pay attention to a lesson before there is any chance in that student remembering or understanding what they are being taught. It is important to not only teach students specific things at schools but also to increase their generally ability to learn.

More specifically, students often took more time after biking on cognitive tests than they did before. Statistically significant accuracy or performance improvements did not always accompany this longer duration, but did in some cases (like the improved error performance on the Go-No Go, for example), and it should be noted that trends were in the direction of improvement. This increase in time is consistent with the explanation that students are less impulsive after biking, and thus are able to take more time on tests. Future academic work on the current biking intervention may include other forms of analysis, including subtractive techniques with the Trail Making Test, item level analysis on the Flanker with respect to previous items, and error proportions pre vs. post on the PEBL Go/No Go. These results are not presented here because they are not intuitively understood, and do not change the general trend of changed cognitive performance after biking (but they do address specific theoretical issues of interest to academics).

The acute Simon results are similar to the Pontifex (2013) results on the flanker. Subjects seem to demonstrate increased interference control after a single episode of biking. Subjects also demonstrate improved accuracy on the TrailMaking test, which could be caused by a number of theoretic cognitive processes. It is interesting that even when subjects are engaging in long term exercise, they still show acute improvements in performance after a single episode of biking. This result of the program suggests that exercise can be used strategically by schools. Students can be scheduled to take the subject they find most difficult or most boring right after an exercise session. This exercise will presumably increase their mood and improve their ability to attend to the subject, and so hopefully improve the student's marks.

Interesting correlations were also found between exercise frequency and intensity measures and mood measures. There appears to be some link between self-reported clarity of feelings, and exercise intensity and frequency that should be specifically investigated in future work. It may be the case that exercise has a positive effect on emotional regulation, or at least a person's understanding of their emotional state. Less than optimal measures of exercise intensity with heart rate monitors made correlation results less interesting than they might have been.

Finding a range of improvements in ADHD subjects is encouraging, but it should be noted that few of the improvements were specific to ADHD subjects. In other words, ADHD subjects did not demonstrate a statistically significant difference from normal subjects in the amount of improvement they made. There is almost certainly some interaction between attention difficulties and at least one other variable going on. For example, Kennedy students with attention problems demonstrated the most obvious trend of improvement on testing after biking, whereas Wilson students with formal ADHD diagnosis seemed to improve the least (with normal subjects being in the middle). There were larger proportions of ADHD students in the biking program at Wilson, a different demographic of students, worse cognitive performance in general, and different reactions to the biking program (the parents at Kennedy very strongly embraced the program and did not attach an ADHD label on it) which could have

led to this trend. Probably these school differences contributed to the results in some fashion; for example, students at both schools might have an increased capacity to attend to school related tasks, but students at Kennedy tried to attend to tasks more. There are numerous other explanations available, and the current intervention cannot tease apart these explanations to demonstrate exactly what was going on.

Subjects with attention problems did, however, demonstrate a trend towards more “normal” brain activity as measured by ERP. The ERP results demonstrate that some kind of change is occurring after biking in neuroelectric activity, a correlate of cognitive processing in the brain. This trend may be analyzed further in more academic work surrounding the biking intervention.

This program demonstrated that it is certainly possible to get Middle School students to bike regularly in the morning. Drop-outs were rare, and even in inclement weather in November, participation was high. Students were also able to maintain levels of sleep while participating in a morning program, according to self-report. However, maintaining levels of outside exercise during a biking program proved to be a difficulty, and future programs should specifically target this issue. Response to an ADHD-centric biking program was positive for both ADHD and non-ADHD students, but note that parent response was so positive toward the program at Kennedy school by parents of non-ADHD students that ADHD student recruitment became difficult.

## **Issues**

Although the intervention was a success at the program level and novel empirical findings emerged from it, there were a number of issues that limit the conclusions that can be made.

### Exercise Issue

One significant issue was that subjects reported less outside exercise when performing the biking intervention. The intervention would almost certainly have been more effective if outside exercise levels were maintained when biking was added. The effects of the intervention were not due to a greater overall duration of exercise, but were instead due to some combination of greater intensity (biking was more intense than the physical activity that they had stopped doing), different timing (biking was at the beginning of the day, dropped outside exercise may have been after school), and perhaps greater frequency of exercise. It is difficult to think of a way of maintaining outside exercise when students engage voluntarily in a biking program other than to encourage this outcome and to have it as a goal of any future program.

### School Differences

There were a number of issues that may have influenced the intervention that were due to differences in the two biking programs at the two schools. Differences in the two programs are potentially problematic, because the two schools acted as controls for one another. Different populations may interact differently with exercise, or with outcome variables that were measured during the study.

The first issue concerns the quality of the biking program. It may be that the biking program was of a higher quality at Kennedy than it was at Wilson. Students at Kennedy had higher average heart rates during exercise, and this statistically significant difference may have also led to differences in results due to this exercise. Correlations between heart rate and outcome variables were shown in the results. This hypothetical quality difference at the schools may also have manifested itself in a higher dropout rate at Wilson. Subjects that dropped out were not available for post measures or in some cases, for the acute measures. Excluding these subjects from the analysis may have biased the results somewhat. It would have been better if mechanisms were put in place that encouraged students who had dropped out of the program to attend all measurement sessions. An intent-to-treat analysis would have measured the performance of subjects even after they had dropped out of the program, and would have provided a very practical measure of biking effectiveness. It is probably the case that the difference in performance between the two schools would have been even greater had post scores for drop-outs been included.

A second issue concerns the populations of the two schools. The Kennedy school outperformed the Wilson school on a number of the measures used in the intervention at pre testing. It is possible that this difference is indicative of a difference in the two populations at the schools. It is easier to start an exercise program with students that are more co-operative, that are more enthusiastic, and that have fewer behavioral issues. These kinds of non-ADHD variables were not specifically tested for, may have been different at the two schools, and may have led to differences in scores between the schools. Also, the generalizability of results may be affected by the school populations.

One final difference between the schools that might have influenced the intervention was that of seasonal effects. Kennedy students performed their biking in cold weather conditions, and a historical hurricane occurred during their intervention. Wilson students biked in much more temperate weather. The weather did not seem to affect the quality of the Kennedy intervention or the intensity of the biking exercise, but it is possible that the determination on the part of Kennedy teachers and RTSG to ensure that students would exercise in adverse conditions may have led to a greater expenditure of effort in comparison to the Wilson intervention.

#### ADHD Issues

There were results in the study that did not reach statistical significance that may have been significant had the number of ADHD subjects in the intervention been larger. Recruitment of ADHD students was an issue that resulted in lower ADHD group numbers than desired at Kennedy. Parents did not want the biking program to have an ADHD label affixed to it, and the need to recruit as many ADHD students as possible was crushed by the great enthusiasm demonstrated by the Kennedy parents of non-ADHD kids. ADHD dropouts were a small issue with Wilson, as a number of ADHD subjects dropped out, reducing the numbers.

Many more formally diagnosed ADHD students were from Wilson than from Kennedy; it was thus difficult to isolate differences between ADHD and non-ADHD students from differences between the two schools, and the programs at the two schools. The issue was much smaller when categories of parent reported attention difficulties were used, and this categorization was thus used in the repeated measures ANOVAs.

The programs were perceived differently at the two schools. The Wilson school biking intervention was acknowledged to be in part an activity to help improve attention. At Kennedy, parents of students were adamant that the intervention not be given any sort of label associated with ADHD. As mentioned above, this may have diluted the message that biking was a potential treatment for ADHD, but it also may have improved the program at Kennedy. Students may feel better about themselves and a biking program if it is not associated with a clinical disorder in any way. The different labels put on the program may have affected the results

Another very interesting potential issue was that the proportion of ADHD students differed in the two schools. The greater the number of ADHD subjects in the intervention, the more difficult it may be to run that intervention. Assuming there were differences in intervention quality at the two schools, this difference may have occurred in part to the different proportion of ADHD to non-ADHD students.

Another ADHD issue was how students were classified with ADHD, and how their ADHD was treated. We used a pre-existing ADHD diagnosis as a criterion for inclusion into the ADHD condition. We also used a more broad definition of ADHD defined by the report of attention difficulties by parents in their children. A third classification scheme revolves around patterns of EEG and ERP results that have been linked with ADHD. Different definitions of ADHD may lead to different results. We also kept track of which students were on medication for ADHD. ADHD medications are typically effective for at least the first year they are taken, but effectiveness has been known to trail off subsequently. All ADHD students taking medication were reported to still be having attention difficulties by their parents, but the medication may still moderate the effect of exercise on these students. In the real world, students with attention difficulties are going to be a heterogeneous mix of formally diagnosed individuals both on and not on medication and non-diagnosed individuals. This intervention was thus an evaluation of biking on a real world population of students with attention difficulties.

There was a lot of inconclusive analysis using different ADHD groupings that was not presented in this document. Medicated vs. unmedicated differences, and attention problem vs. formal diagnosis differences were thus not covered in this report. The numbers of students in the various categories (medicated ADHD, unmedicated ADHD, attention difficulty without diagnosis) were not large, and so the probability of finding statistically significant differences between these particular groups was smaller than it could have been if larger groups were used.

One final note concerning ADHD students: differences in quality of teachers at a school can influence ADHD outcomes, especially in the cognitive domain. If academic programs at a certain school are more suitable for students who suffer attention difficulties, or if such students receive more support at a certain school, then this difference will affect the results of any biking program tailored toward such students.

### Measurement Issues

There were a couple of issues surrounding the measures that were used in the study. The first issue concerns questionnaire responses. These responses were all potentially subject to self-report bias. For example, self-report is susceptible to social desirability bias, where responders put in an answer that will be perceived favorably by others. This bias can also interact with the particular group a person is in.

For example, ADHD students at Wilson, aware that the biking they were doing was intended to help them, may have been more motivated to report an attention benefit of the exercise than non-ADHD students.

A second issue concerns the more academic measures that were used. These measures had the benefit of being objective, with a known degree of reliability and validity in most cases (issues surrounding specific measures are covered in the measures section). One general point regarding these measures is that they are often more broadly labeled than is warranted. For example, the Flanker and the Simon tasks are labeled tests of attention, but they correspond to specific facets in a particular theory of attention. It is possible for attention to improve, but not in the specific areas that these tests measure. Also, some of the measures used were computerized versions of common cognitive tests. Although these tests are analogous to common measures, the computerized presentation may have changed the test in unforeseen ways.

One final measurement issue is the duration of measure. The current biking intervention lasted one month, and subjects were measured no longer than one month after the intervention. Long term outcomes of the biking intervention were not measured.

### **Future Directions**

Measuring the long term outcomes of biking interventions would be a good idea. Assuming exercise has an acute benefit to attention as demonstrated in this intervention, it would be beneficial not only for a student to receive this attention boost for the duration of a one month program, but also after the exercise program is over (by using exercise as a clinical tool in their everyday life). Similarly, assuming there are long term benefits to exercise for attention, it would be useful to determine how long these benefits last after exercise stops, how much benefit accrues over longer durations, and how difficult it is to maintain clinically beneficial exercise over longer time frames.

In the future, it would be better to examine the specific strengths and weaknesses of a biking in comparison to other modes of exercise in a program aimed at helping attention difficulties. The results of this study point to a number of strengths and weaknesses of biking in comparison to other forms of exercise:

1. Biking done in a group is intensity variant, and pace invariant: all participants must maintain a pace to stay together, and exercise is more intense for those who are in worse shape. Higher intensity exercise is better to a point, but exercise at too high an intensity can lead to anaerobic intensities. Such exercise is much less pleasant, and increases the chances of program drop out. It would be better to have biking exercise that was intensity invariant, with subjects biking to maintain their heart rate in a target zone. Such biking would have to be done individual, since students could not bike in a group when exercising at different paces.
2. Biking and exercise compliance: biking outdoors may be a form of exercise that is easier to maintain than other forms of exercise. No comparisons were made between biking and other forms of exercise in this intervention, but program drop-out rates appear to be very

- low, and were due mostly to difficulties in parents transporting their children to school. Drop-out rates were 0% at Kennedy when students were biking in inclement weather.
3. Biking season and location: Kennedy school biked in inclement weather, but had more scenic biking routes. These two program components may have canceled one another out. Biking is potentially better than other forms of exercise because it allows for outdoor exercise in nature. It would be better in the future to examine biking during a warm weather season in scenic outdoor conditions, since this kind of biking might lead to improved results, especially in the mood domain, and compare this exercise to equally intense exercise indoors (stationary biking). During the current intervention, any hypothetical improvement of mood due to exercise may have been canceled out by a hypothetical decrease in mood due to seasonal effects. Mood measures are potentially subject to calendar effects, where students may have worse moods nearer to exam times. Also, it would be better to have pre-testing done in the middle of the school year, as well as post testing; this would control for general school stress level better, since students are less likely to be stressed about school work at the beginning of the year, when nothing is due.

#### Summary Suggestions for Future Program Design

1. Don't label biking programs as ADHD-centric
2. Maintain a small ADHD population in a larger biking group rather than having a large ADHD group
3. Bike during seasonable weather and in scenic locales for best results, but do not assume that students will not bike in inclement weather
4. Use biking group as a training ground for self-paced, self-directed exercise, don't conceptualize it as the intervention
5. Focus exercise initially on fun and a demonstration of competency and some minor benefits (have participants compare their mood before exercise with their mood after a minute of exercise, for example). Only when participants believe in the benefits of exercise, believe they can successfully perform exercise, and like the exercise that they are doing should a program be exercise intensity focused.

#### Summary Suggestions for Future Research

1. Compare different biking groups within the same school, or enlarge the study to include school as a random variable and include a number of schools in every condition
2. Compare different ADHD proportions in biking groups to determine which is best
3. Compare outdoor biking to stationary biking indoors
4. Measure the use of "as needed" biking, where subjects can use a stationary bike in class when they find attention wandering, or they are having difficulty self-regulating
5. Measure the long term effects of biking (what is result of biking over a number of years?)

## Appendices

- A. Questionnaire (Post Version, includes Trait-Meta Mood, Connor's Screening, and Survey Evaluation)

Subject ID: \_\_\_\_\_

Date: \_\_\_\_\_

**Questionnaire**

1. Approximately how many hours do you watch T.V., play video games, go on the internet or use a computer on the usual school day (think about how many hours before school, during school and after school)? You can use 0 or .5 (30 minutes) or 1 or more.

\_\_\_\_\_ hours before school \_\_\_\_\_ hours during school \_\_\_\_\_ hours after school

How many hours do you watch T.V., play video games, go on the internet or use a computer on the weekend?

\_\_\_\_\_ hours over Saturday and Sunday

2. Approximately how many hours are you asleep at night on a usual school night?  
I usually fall asleep at \_\_\_\_\_(time) and wake up at \_\_\_\_\_(time) for school

How many hours are you asleep at night on a usual weekend night?  
I usually fall asleep at \_\_\_\_\_(time) and wake up at \_\_\_\_\_(time) on the weekend.

3. Do you do any other physical activities besides this biking program? (any sports or games you play) Write them below.

---

---

4. Approximately how many hours per week are you doing the physical activities listed above? (Think if you have practice on certain days, or usually play outside of practice on your own.) Remember you can use 0, or .5 (30 minutes) or 1 or more.

\_\_\_\_\_ hours Monday \_\_\_\_\_ hours Tuesday \_\_\_\_\_ hours Wednesday \_\_\_\_\_ hours Thursday  
\_\_\_\_\_ hours Friday \_\_\_\_\_ hours Saturday \_\_\_\_\_ hours Sunday

5. Put an X in the box that best describes your answer to the following questions

	Not at all	A little	Some	Quite a bit	A lot
How much do you enjoy exercise?					
How well do you get along with your classmates?					
When you wake up in the morning, how awake do you feel?					

Tell us about your experience in this bike program.

Did you have fun?

Would you tell a friend to do this program?

Would you do it again?

What did you like the most?

If we do this again, what should we change? How can we make it better?

How true are the following statements for you? Circle your response.

6. I often think about my feelings.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

7. It's usually a waste of time to think about your feelings.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

8. I believe you should do whatever your feelings tell you to do.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

9. I pay a lot of attention to how I feel.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

10. The best way to handle my feelings is to just go ahead and feel whatever I'm feeling.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

11. I believe it's good for you to go ahead and feel whatever you feel.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

12. My feelings help me decide how to act.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

13. I almost always know how I'm feeling.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

14. I usually know how I feel about things.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

15. I am comfortable with my feelings

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

16. I am usually very clear about my feelings (I usually know which feeling I am having).

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

17. I am usually confused about how I feel.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

18. If I find myself getting mad, I try to calm myself down.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

19. I try to think about good things no matter how bad I feel.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

20. When I become upset, I think about all of the good things in my life.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

21. No matter how bad I feel, I try to think about good things.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

22. Even though I am sometimes happy, I mostly think bad things are going to happen to me.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

23. When I am happy, I realize how silly most of my worries are.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

24. Even though I am sometimes sad, I usually think good things will happen to me.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

25. When I'm sad I can't help thinking about bad things.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

26. When I am upset, I realize that the good things in my life aren't really good after all.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

27. When I'm in a bad mood, I think lots of bad things will happen to me.

Not at all True	Hardly Ever True	Sometimes True	Often True	Always True
-----------------	------------------	----------------	------------	-------------

One More Page to Go!!!

You're Almost Done 😊

For each description of actions or behaviors in the left column, put an X in one of the columns to the right that best indicates the degree to which you have acted in the past week.

Actions	Not at all (0)	Just a little (1)	Pretty much (2)	Very much (3)
Restless and overactive				
Excitable, impulsive				
Disturbs other children				
Fails to finish things.				
Short attention span				
Constantly fidgeting				
Inattentive, easily distracted				
Demands must be met immediately				
Easily frustrated				
Mood changes quickly and drastically				

B. Additional Graphs of Test Results

C. QEEG Results

D. ERP Attention Groupings

## E. References

- Archer, T., & Kostrzewa, R. M. (2012). Physical exercise alleviates ADHD symptoms: regional deficits and development trajectory. *Neurotoxicity research*, 21(2), 195-209. doi:10.1007/s12640-011-9260-0
- Best, J. R. (2010). Effects of physical activity on children's executive function: Contributions of experimental research on aerobic exercise. *Developmental Review*, 30(4), 331-351. doi:10.1016/j.dr.2010.08.001
- Blumenthal, J. a. (2011). New frontiers in cardiovascular behavioral medicine: comparative effectiveness of exercise and medication in treating depression. *Cleveland Clinic journal of medicine*, 78 Suppl 1, S35-43. doi:10.3949/ccjm.78.s1.06
- CDC (2010). The Association Between School-Based Physical Activity , Including Physical Education , and Academic Performance. *Education*, (April).
- Chalder, M., Wiles, N. J., Campbell, J., Hollinghurst, S. P., Haase, a. M., Taylor, a. H., Fox, K. R., et al. (2012). Facilitated physical activity as a treatment for depressed adults: randomised controlled trial. *Bmj*, 344(jun06 1), e2758-e2758. doi:10.1136/bmj.e2758
- Chang, Y. K., Labban, J. D., Gapin, J. I., & Etnier, J. L. (2012). The effects of acute exercise on cognitive performance: a meta-analysis. *Brain research*, 1453(250), 87-101. Elsevier B.V. doi:10.1016/j.brainres.2012.02.068
- De Pauw, S.S., Mervielde, I. (2011). The role of temperament and personality in problem behaviors of children with ADHD. *J Abnorm Child Psychol*, 39, 277-291.
- Dutton, D. G. and Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology*, 30, 510-517.
- Eriksen, C.W., Eriksen, B.A. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Percept Psychophys*, 5, 249-63.
- Gapin, J., Etnier, J.L. (2010). The relationship between physical activity and executive function performance in children with attention- deficit hyperactivity disorder. *J Sport Exerc Psychol*, 32, 753-763
- Gapin, J., Labban, J.D., Etnier, J.L. (2011). The effects of physical activity on attention deficit hyperactivity disorder symptoms: the evidence. *Prev Med*, 52, S70-S74.
- Halperin, J. M., & Healey, D. M. (2011). The influences of environmental enrichment, cognitive enhancement, and physical exercise on brain development: can we alter the developmental trajectory of ADHD? *Neuroscience and biobehavioral reviews*, 35(3), 621-34. Elsevier Ltd. doi:10.1016/j.neubiorev.2010.07.006
- Hill, L., Williams, J. H. G., Aucott, L., Milne, J., Thomson, J., Greig, J., Munro, V., et al. (2010). Exercising attention within the classroom. *Developmental medicine and child neurology*, 929-934. doi:10.1111/j.1469-8749.2010.03661.x

- Hopkins, M. E., Davis, F. C., Vantighem, M. R., Whalen, P. J., & Bucci, D. J. (2012). Differential effects of acute and regular physical exercise on cognition and affect. *Neuroscience*, *215*, 59-68. IBRO. doi:10.1016/j.neuroscience.2012.04.056
- Kerns, J.G., Cohen, J.D., MacDonald, A.W.I., Cho, R.Y., Stenger, V.A., Carter, C.S.(2004).Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023-6.
- Laurent, J., Catanzaro, S. J., Rudolph, K. D., Joiner, T. E., Potter, K. I., Lambert, S., Osborne, L., et al. (1999). A Measure of Positive and Negative Affect for Children : Scale Development and Preliminary Validation. *Psychological Assessment*, *11*(3), 326-338.
- Mullane, J. C., Corkum, P. V., Klein, R. M., & McLaughlin, E. (2009). Interference control in children with and without ADHD: a systematic review of Flanker and Simon task performance. *Child neuropsychology : a journal on normal and abnormal development in childhood and adolescence*, *15*(4), 321-42. doi:10.1080/09297040802348028
- Pontifex, M. B., Saliba, B. J., Raine, L. B., Picchiatti, D. L., & Hillman, C. H. (2013). Exercise Improves Behavioral, Neurocognitive, and Scholastic Performance in Children with Attention-Deficit/Hyperactivity Disorder. *The Journal of pediatrics*, *162*(3), 543-51 doi:10.1016/j.jpeds.2012.08.036
- Rockhill, C.M., & Greener, S.H. (1999). *Development of the Trait Meta-Mood Scale for elementary school children*. Poster presented at the biennial meeting of the Society for Research in Child Development, Albuquerque, NM.
- Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, *20*, 343-350.
- Sibold, J. S., & Berg, K. M. (2010). Mood Enhancement Persists for Up To 12 Hours Following Aerobic Exercise: a Pilot Study 1. *Perceptual and Motor Skills*, *111*(2), 333-342. doi:10.2466/02.06.13.15.PMS.111.5.333-342
- Smith, P. J., Blumenthal, J. a, Hoffman, B. M., Cooper, H., Strauman, T. a, Welsh-Bohmer, K., Browndyke, J. N., et al. (2010). Aerobic exercise and neurocognitive performance: a meta-analytic review of randomized controlled trials. *Psychosomatic medicine*, *72*(3), 239-52. doi:10.1097/PSY.0b013e3181d14633
- Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and healthy adults. *Neuropsychology*, *11*, 138-146.