The Effects of Exercise on Cognitive Performance

Investigators: Matthew Finn, Alisionna Iannacchione, Kyle Smith

Advisors: Kajal Claypool, Mark Claypool, Adam Lammert

Worcester Polytechnic Institute

Abstract

Previous studies have connected exercise to increased cognitive performance. Specifically, one single instance of moderate exercise has been shown to increase working memory (Hogan, et al., 2013). However, there is a lack of information and data on how college students may benefit cognitively from a session of exercise. The goal of our project is to explore relationships between exercise and cognition. We would like to correlate the exercise faculties of physical exertion and movement with the cognitive faculties of attention and memory. We performed our experiment by giving cognitive tests to two groups: students participating in gym classes and students resting while watching a show of their choice. We measured physical exertion with the activity trackers Fitbit Inspire HR and Axivity AX3 and measured cognitive ability using the Flanker and Corsi tests. Based on our experiment, with 32 participants in the control group and 30 in the exercise group, we found that both groups showed a significant improvement on the attention test while only the exercise group showed a significant improvement on the memory test. We were also able to distinguish the exercise group from the control group with significant differences in both heart rate and movement, showing that the exercise group exerted much more effort during their activity sessions. There was not a statistically significant correlation between heart rate or movement with increased cognitive performance.

Table of Contents

1 Introduction

Exercise is known to benefit many aspects of general health. A 2005 review of the mental and physical health benefits of exercise concluded that several populations and age groups all benefited from physical exercise in the decreased likeliness of obesity, cardiovascular disease, cancer and sexual dysfunction (Penedo, 2005). Furthermore, exercise was shown to decrease the chances of depression and increase general mood states.

Previous studies have also shown that exercise is linked to positive effects on cognitive performance. Many of these effects are seen in older populations since although aging can deteriorate the brain, brain scans have shown that there is a link between higher cardiovascular health achieved from working out and less brain tissue loss with age (Colcombe, et al., 2003). Other studies have shown that older populations are not the only ones that can benefit cognitively through exercise. For example, according to a study conducted by the psychology department at Stanford University measuring exercise and cognitive performance to see if age had an effect on the benefits of exercise, results showed that a single instance of exercise had positive effects on cognitive performance regardless of age (Hogan, et al., 2013). Many studies have shown that being active versus inactive actually improves results in a variety of cognitive tasks (Hogan & Kiefer, 2013; Kamijo, 2009). In a meta-analysis conducted by the Department of Kinesiology at the University of North Carolina, researchers compared 79 studies measuring the effects of exercise on cognition and found that there were small but positive effects in a large majority of the time (Chang, et al., 2012). The effects varied based on factors like the type and duration of the exercise and cognitive tasks used to measure cognitive performance.

For college students, an increase in physical activity may provide a benefit in classroom performance and subsequently in their careers. More data would be helpful in determining whether or not acute sessions of exercise would immediately assist college-aged people in a cognitive manner. This could help students in deciding whether or not to exercise, what kind of exercise to perform, and when to exercise. Different kinds of exercise may lead to different benefits in cognitive performance; both aerobic and anaerobic physical activity are popular among people who exercise. The American College of Sports Medicine (ACSM) defines anaerobic exercise as “intense physical activity of very short duration, fueled by the energy sources within the contracting muscles and independent of the use of inhaled oxygen as an energy source” (American College of Sports Medicine, 2013). An example of this is high intensity interval training, where exercisers exert as much energy as possible and then rest for an extended period until they are ready for another session of intense exertion. The ACSM defines aerobic exercise as “any activity that uses large muscle groups, can be maintained continuously and is rhythmic in nature” (Wahid, 2016). An example of aerobic exercise is walking or jogging. A 1999 six-month study examined 124 adults randomly assigned to either aerobic or anaerobic exercise and discovered that the aerobic group experienced obvious improvement while performing executive control tasks, and the group that trained anaerobically improved much less (Kramer, 1999). It may be the case that steady-state cardiovascular exercise is more helpful than most forms of exhaustive exercise when it comes to cognitive performance. Another paper from 2002 analyzed three groups of experiments to assess whether or not adults benefitted cognitively from acute sessions of exercise. The review concluded that up to 60 minutes of aerobic exercise without exhaustion improved some aspects of information processing. Longer exercise that led to exhaustion and/or dehydration harmed memory and information processing (Tomporowski, 2002). Further investigation is needed to conclude whether or not intense anaerobic exercise may enhance cognitive performance.

Despite the many potential advantages of exercise, many college students do not fully realize the cognitive benefits they may receive from physical activity. Data collected from the CDC (Centers for Disease Control and Prevention), showed a decline in the amount of regular physical activity from students in high school (65%) to students in college (38%) (Calestine; Bopp; Bopp, C. M.;Papalia, (2017) taken from the Centers for Disease Control and Prevention, 2016).

The goal of our project is to study the effects of acute physical activity in college students on cognitive performance by examining the immediate changes in cognitive ability before and after physical exertion. We developed two hypotheses to test during experimentation, based on background investigation and research. H1: We expect that acute physical activity will improve short-term cognitive performance more than a period of inactivity. This hypothesis is supported by the large majority of the research into exercises' benefits. H2: We expect that the subjects who exercise more regularly will experience greater improvements in cognitive performance after exercise than the subjects who exercise regularly. This may be the case because participants that exercise regularly have adapted to its benefits and may not have a dramatic increase in cognitive ability compared to irregular exercisers.

College students (ages 17-26) from Worcester Polytechnic Institute were studied by performing a normal session of their gym classes and completing the same cognitive testing before and after their session of exercise (30 exercise participants). There was a control group that performed the same activities as the exercise group but instead of participating in a gym class, the students sat and watched a television show (32 control participants). The task was completed using cognitive tests to measure attention and memory along with activity trackers to measure physical exertion during exercise. The cognitive test that the subjects participated in before and after their activity/inactivity tested the cognitive aspects of attention and memory. The activity trackers gathered heart rate data and accelerometer data. The novelties of this study are testing the theory of improved cognition from exercise on college students and observing the cognitive performance of college students that have reached exhaustion.

The results of the experiment showed that there was not a statistically significant difference in the improvement of the cognitive tests between the exercise group and the control group. There was also not a statistically significant correlation between exercise habits and improvement in cognitive testing. However, the exercise group improved significantly in the Corsi test after they exercised, while the control group did not improve significantly after watching a show. There was not a statistically significant correlation between heart rate or movement with increased cognitive performance.

The rest of this paper is organized as follows: Chapter 2 provides background research into topics in the space of our experiment. Chapter 3 explains the experimental methodology. Chapter 4 reports the results of the experiment. Chapter 5 is a discussion about the results of the experiment.

2 Background

The purpose of the following background investigation is to determine two essential aspects of the experiment: 1) the cognitive tests that will be completed by the participants before and after physical activity and 2) the wearable devices that will be used by the participant during physical activity. There are a wide variety of methods to measure cognition: some tests examine multiple aspects of cognitive performance (memory, attention, mental agility, etc.) while others measure just one or two. This chapter describes different kinds of tests and different aspects of cognitive performance to determine what is most effective for our experiment. Similarly, there are many ways to measure physical exertion. Some devices measure heart rate, speed, steps, acceleration, and other aspects of physical exertion, usually with a wearable device. The convenience, amount of data available, and the number of physical measures will be examined across multiple exercise devices to select the equipment that will best fit the need of the experiment.

2.1 Introduction to Cognitive Tests

Cognitive tests are a developed and widely accepted way clinicians, psychologists, and scientists assess the underlying state of cognitive functioning either to diagnose mental illness, such as ADHD or Alzheimer's, or test on situational bases to retrieve data on how cognitive functions can fluctuate for healthy individuals based on other factors (Committee on Psychological Testing, 2015). “The term *cognitive functioning* encompasses a variety of skills and abilities, including intellectual capacity, attention and concentration, processing speed, language and communication, visual-spatial abilities, and memory” (Committee on

Psychological Testing, 2015). These skills and abilities have been measured using a wide variety of tests in the clinical and research setting.

2.1.2 Current Standards in Cognitive Testing

Quite often, previously used testing measures are replicated or adapted to test cognitive functions for a variety of situations. In a clinical setting, the test may be a measure to confirm already existing symptoms presented by a patient. Although patients often self report their cognitive state, it may not be as accurate as assessing via a test and comparing with trends shown using a large population of people with similar mental and physical diagnosis (Committee on Psychological Testing, 2015). In the psychological setting, testing a cognitive function as a result of a changing variable can be difficult, the changes are slight, and controlling for confounders is more crucial to the findings.

Reliability in the space of psychological and biological research is most often on the basis of test-retest reliability, concurrent validity, and replication of results. For example, a longitudinal study was conducted to test the reliability of the Gibson Test of Cognitive Skills and tested the same participants multiple times as well as correlated the findings with other reputable studies. The Gibson Test was found to be a valid and reliable way to measure cognitive functions like memory, processing and reasoning (Moore, et al., 2018). It is important for the purposes of our experiment to choose a test that has shown reliability in past research.

2.1.3 Pen-and-Paper or Computer-based Laboratory Testing

While many of the clinical tests for cognition are still done using pen and paper, there is a recent emergence of computerized testing in some clinical settings and more preferably in the area of psychological research (Committee on Psychological Testing, 2015). Many wonder how equivalent these two methods of testing are as far as reliable results and how they compare when deciding which to use. Older studies from the late 1980s to the mid-1990s suggest that people tend to read and comprehend information at slower speeds if reading from a screen than from paper (Noyes, et al., 2008). A 1993 meta-analysis examined if there was a "medium effect" for cognitive testing: if the medium or means used to administer the test made a difference in the results. This analysis focused on studies looking at power tests and speed tests, finding that power tests (tests looking at correctness not time constrained but still timed), did not exhibit this medium affect, whereas a medium affect was found with tests measuring speed (Mead & Drasgow, 1993). The new influx of technology might render this to be much different today. For example, a 2017 study compared computerized cognitive testing versus pen and paper to measure early signs of Alzheimer's disease compared the National Institute of Health's Cognitive Battery Test (NIH Toolbox) and the Cogstate C-3 cognitive test (both cognitive tests that can be administered on a computer or tablet) to Preclinical Alzheimer Cognitive Composite (PACC), the current standard pen-and-paper assessment. Their results showed the NIH Toolbox was able to detect cognitive functions closest to the PACC test and for the C-3's group of tests, some compared well to the PACC and some did not (Buckley, et al. 2017). This shows that reliable computerized testing can render the same results as pen-paper. While both methods of administering cognitive tests have benefits and limitations, the new popularized method of computerized tests for cognitive functions can be a more useful option for this study as it allows for easy data collection and coding responses.

2.1.4 Recent Advances in Unobtrusive Monitoring of Cognitive State

Recent research has investigated ways to monitor cognitive abilities without using specialized tests that subjects have to attend at all. New studies are showing that facial expressions and speech can be a way to measure cognition in a way that is not obtrusive (Lammert, et al., 2017; Sloboda, et al., 2018). These methods have been supported in military environments due to the physical and mental fatigue they experience as well as not having time to focus on conducting studies. There is promise that monitoring through speech can be applied outwards to other populations. A study looking at the cognitive effects of mild traumatic brain injury (MTBI) created models to estimate the processing speed index of normal subjects as well as a subject with MTBI. When comparing results from traditional cognitive tests and facial and speech markers measured in audio and video from the subjects showed high correlations (Lammert, et al., 2017). Another study looking at speech in measuring cognition looked at speech in the context of fatigue and used fatigue and load to help measure cognitive performance (Sloboda, et al., 2018). They used markers to distinguish fatigue and load during a working memory task: "Speech onset time, speaking rate, and vocal tract coordination features show strong potential for speech-based fatigue estimation" (Sloboda, et al., 2018). These new forms of cognitive testing seem promising due to the exploration of these ideas giving similar results to standard tests but they still need more reliability. Cognitive testing in a way that is not obtrusive through using methods of speech and facial expressions in upcoming research could be a novel way to add to this data.

2.1.5 Summary

Table 1: Broad Overview of Cognitive Tests

What is Assessed	Means of Measuring	Example
Memory	Recognition Tasks Recall Tasks Word Fragment Tasks Memory Functioning Questionnaire N-back test Visual and Verbal Memory Test Gibson Test of Cognitive Skills (Sternberg & Sternberg, 2016)	N-back Test: One version consists of being told a list of letters or numbers. In a 3-back design, subjects would be asked to recite the 3rd letter previously spoken at the appropriate time (Farnsworth, 2016).
Reasoning	Simon Task Gibson Test of Cognitive Skills Abstract Reasoning Test Deductive Reasoning Test Inductive Reasoning Test Numerical Reasoning Test Verbal Reasoning Test Flanker Task (Sternberg & Sternberg, 2016)	Abstract Reasoning Test: One variation includes questions looking at patterns of figures by shapes, sizes and directions and deciding which figure would come next in the sequence or being asked to fill in a portion of a picture based on the pieces already shown (123 Test, 2019).
Perception	Motor-Free Visual Perception Test Music-In-Noise Task Motion Direction Discrimination Task (measures Perceptual Learning) (Sternberg & Sternberg, 2016)	Music-In-Noise Task: Subjects are asked to listen to two different melodies and asked to distinguish if they are the same. One is played alone and the other is played over a masking sound (Coffey, et al., 2019).
Attention	Stroop Task Trail Making Test D2 Test Target Detection Task Gibson Test of Cognitive Skills Psychomotor Vigilance Task Flanker Task (Sternberg & Sternberg, 2016)	The Stroop Test: Subjects are shown colors and are timed at how long it takes to recite them. They are shown a color word in the color that the word is describing and asked to recite the words while timed. They are then shown color words that are not colored in the color the word is describing. They are timed on how long it takes to recite (Scarpina & Tagini, 2017).

Language	Verbal Fluency Test Cognitive Linguistic Quick Test (Sternberg & Sternberg, 2016)	Verbal Fluency Test: Subjects are asked to speak as many words as they can in a timed period (Shao, et al., 2014).
Executive Functioning	Flanker Task Go/No go Task Stop Signal Tasks Gibson Test of Cognitive Skills Simon Task (Sternberg & Sternberg, 2016)	Flanker Task: Subjects are shown a variety of arrows. In each trial, one arrow is highlighted and the subject is asked to determine what direction the arrow is moving. While some trials include the highlighted arrow pointing in the same direction as the other arrows, other trials include arrows pointing in other directions in the same line (incongruent stimuli) and some trials include neutral symbols with no direction (Eriksen & Eriksen, 1974).
Action	Purdue Pegboard Task Psychomotor Vigilance Task (Sternberg & Sternberg, 2016)	Purdue Pegboard Task: Subjects will be presented with a board with two vertical rows of holes. They will be timed for 30 seconds to take pegs and put them into the right line of holes one at a time with their right hand. They will be asked to repeat with their left hand on the left side for 30 seconds. They will be asked to use both hands to do both rows at the same time for 30 seconds. The last task involves inserting a peg into the right row with their right hand and subsequently covering the peg with a washer with their left, a colander with their right, and another washer with their left hand. This task is given 60 seconds (Rehabilitation Measures Database, 2014).

Table 1 lays out a framework of what can be measured using cognitive tests and which tests cover each area of interest. The leftmost column, “What is Assessed,” represents different cognitive faculties that can be measured through cognitive testing. To the right of that, “Means of Measuring” lists popular cognitive tests that are used to measure each area of cognition. The rightmost column, “Example,” explains how one test from each list works. For example, the

N-back test can assess memory and one variation of this test is described in the right column. Many of these tests can be shown to measure multiple fields of cognition. For example, the Gibson test of Cognitive skills has many aspects that can get at measuring reasoning, perception, and memory (Moore, et al., 2018).

Table 2: Access and Administration of Select Cognitive Tests

Cognitive Test	What is Measured	Pen-Paper vs Computerized	How Long to Administer	Cost
N-back	Memory	computerized	1-40 minutes	$30 (Dual N-back Pro)
Simon Task	Reasoning Executive Functions	computerized	2+ minutes	Free Online Version
Motion Direction Discrimination (MDD) Task	Perception Learning	either	~20 minutes	$75-175
Stroop Task	Attention	either	1-10 minutes	$120
Verbal Fluency	Language	either	1-5 minutes	Free Online Version
Flanker Task	Executive Function Attention Reasoning	computerized	~3 minutes	Free Online Version
Psychomotor Vigilance Task (PVT)	Action Attention	computerized	< 10 minutes	Free Online Version

Table 2 provides more details on a cognitive test that measures one or more of the seven cognitive functions laid out in Table 1. In this table, the leftmost column lists the test and to the right of that is what is measured, how the test can be administered (be that pen and paper or computerized or both), how long it takes to administer, and how much it costs to buy an official version. This table is contextualized to compare potential tests that could be used to retrieve data for this project as well as provide a reference for future studies when assessing which cognitive test is best to use.

Table 3: Past Research using Cognitive Tests to Access Changes Before and After Exercise

Cognitive Test	**Past Research**
N-back	Independent of age, aerobic exercise resulted in faster reaction times in a 2-back test compared to control participants.(Hogan & Carstensen, 2013).
Simon Task	A study looking at aerobic exercise on cognitive control for participants ages 23-63 found that exercise did not improve cognitive control measured by the Simon Task (Joyce, et al., 2014).
Motion Direction Discrimination (MDD) Task	Contrary to previous results, in a study of 27 adult participants, ages 18-60, over a 5 day period, aerobic exercise actually reduced perceptual learning compared to the resting condition (Connell, et al., 2018).
Stroop Task	A study measuring 20 participants ranging from 19-24 years old found, one acute bout of moderate exercise improved the effect of Stroop interference(Yanagisawaa & Dan, 2010).
Verbal Fluency	A study measuring exercise as a form of rehabilitation for patients with chronic pulmonary heart disease found that exercise (Emery, et al., 1998).
Flanker Task	A study measuring changes in cognitive performance due to acute steady-state exercise found that this type of exercise improved reaction times in a Flanker Task (Davranche, et al., 2019).
Psychomotor Vigilance Task (PVT)	A study measuring aerobic fitness in young adults found that young adults who have higher fitness levels have slower reaction times in the PVT than lower fitness levels (Ciria, et al., 2017).

Table 3 provides examples of specific cognitive tests being used in past research involving exercise. The leftmost column shows which test has been used in past research involving exercise and cognition (the tests are the same as shown in Table 2). The right column summarizes results from a research study using that cognitive test to measure a cognitive faculty changing due to exercise.

2.2 Introduction to Devices Used to Measure Physical Activity

Before the recent influx of portable technology, measuring activity was not an exact science. People could exercise for health benefits, but measuring and logging this data was very difficult and not possible for the average person. In 1981, Polar created the first commercially available activity tracker for use by sports professionals and athletes ("Olympic Medical Institute Validates Polar RS800 Running Computer And Training System", 2014). This device was only meant to measure heart rate, but other companies quickly realized that there are other things to be measured, thus creating an entirely new industry. In March 2008, the government became involved and created their own activity tracker for Americans to use during the National President's Challenge, where President Bush called on Americans to exercise to help prevent obesity, and created an activity logger program on the White House website for people to use to see how well they are exercising ("President Bush Discusses National President's Challenge", 2008). In the modern-day, an activity tracker is any wearable device that can measure physical activity. What was created in 2008 was not an activity tracker in the modern sense of the phrase, but it allowed everyday people to set their goals and see how close they are to meeting them.

In the past few years, activity tracking devices have become nearly ubiquitous. Most modern smartphones today have methods of either logging activity or using sensors to measure activity or biological functions, such as heart rate. Wearable activity trackers, such as the Fitbit or Apple Watch, are increasing rapidly on the market. According to statista.com, a website for the pooling of statistics, the global revenue of activity trackers is predicted to be $33.78 billion in 2019, compared to a measured $16.07 billion in 2016 (Shanhong, Liu, 2019). This research study will choose a device to measure if subjects are reaching adequate activity levels and then use a cognitive test defined in the previous sections to measure whether such activity can increase cognitive performance.

2.2.1 Types of Activity Trackers

Activity trackers today come in many different forms. The most common one is the wearable kind, often watches, but smartphones can also measure activity, and some studies utilize an EKG-like device to measure heart rate. Watches are most commonly used due to: 1) their ease of use, because they are meant to be used by the average consumer; 2) their power, because such a small device can still hold all the necessary hardware; and 3) the convenience, because it's as easy as putting on a watch. Some notable examples of wearable fitness trackers include the Fitbit line of products, the Apple Watch, and the Samsung Gear line of products.

Most modern smartphones also have the ability to measure activity levels. Samsung-brand phones come installed with the app Samsung Health, which can be used to store data, including heart rate, an activity log, blood pressure, and blood oxygen level. Apple-brand phones come installed with the app Health, which can measure sleep, activity, and also syncs with the Apple Watch to record steps and heart rate directly to the app. Some wearable activity trackers come with their own apps to record data from the device, such as the Fitbit app.

2.2.2 Metrics

Choosing a device for a study of this kind can prove difficult. There are many things to consider, like the price, convenience of the device, and what it measures. Most importantly, the data gathered from these devices must be trustworthy. Most of these devices "process" the data, meaning it is impossible to get the raw sensor data, which may be desirable in some cases. Processed data may be fine, but some research must be done into what the software does to the data before it is ready for use. Some of the most common sensors include gyroscope and accelerometer for measuring movement, and a heartbeat sensor. Table 4 shows some sample devices and the sensors they have and what they measure.

Table 4: Sample Devices with Sensors and Measurements

Device	Device Type	Accelerometer	Gyroscope	Heartbeat	Pulse Oximeter	Cost
Fitbit Inspire HR	Watch	Triaxial	No	Yes	No	$99.95
Axivity AX3	Watch	Triaxial	No	No	No	$134.31
Faros 360	Misc	Triaxial	No	Yes (heart rate variability)	No	$3494.30
Samsung Gear S2 Classic	Watch	Triaxial	Yes	Yes	No	$299.00
Polar Vantage V Titan	Watch	Triaxial	Yes	Yes	No	$599.95
Apple iPhone X	Phone	Triaxial	Yes	No	No	$730.00
Samsung Galaxy S10+	Phone	Triaxial	Yes	Yes	Yes	$999.99
ActiGraph wGT3X-BT	Watch	Yes	No	No*	No	$225.00

* Can connect a Polar H10 or H7 Bluetooth Heart Rate Monitor

An accelerometer is a sensor that measures the strength of all acceleration forces acting on it, including the force of gravity. Most activity trackers will have one since it is the most common sensor for measuring movement. A triaxial accelerometer is a type of accelerometer that measures acceleration in three directions: front and back, side to side, and up and down. A gyroscope is used to measure orientation and angular velocity. A heartbeat sensor is also very common among activity trackers since heartbeat is a very good measure of physical exertion. Finally, the pulse oximeter is less common on activity trackers but commonly used in medical settings where a person's blood oxygen saturation is important.

If, for example, the Galaxy S10+ is to be used in a study of this kind, it has the most sensors and measures the most aspects of activity, but has two significant drawbacks: the logistics of using such a device, and the price. People who exercise usually wear loose-fitting clothing that either has no pockets or shallow pockets that are incapable of holding a device this size (the screen has a diagonal length of 6.4 inches and a weight of 175g, larger and heavier than all the devices on this page and most other phones). Additionally, the heartbeat sensor does not measure constantly. Instead, it has a small sensor on the back of the phone which will measure heartbeat if a finger is placed on it, and only measures for a few seconds at a time. It is also the second most expensive device on this table, with a price tag of $999.99, so acquiring multiple devices for a study is not feasible.

2.3 Previous Studies

To measure physical activity, A.M. Khan used the Axivity device and determined that the Axivity device had an "accuracy" of approximately 92% (Khan, 2012). The Axivity was also used in a large study that analyzed 96,220 UK-Biobank participants to see if there is a correlation between physical activity and cognitive performance in adults (Willetts, et al., 2018). Kamijo used an unspecified Polar watch in a study, which is made by the same company as the Polar Vantage V Titan specified in the previous table (Kamijo et al., 2009). Polar has been around since 1977 and has been used in this study and others, so this would also be a reliable device

2.4 Summary

Cognitive testing can be useful in measuring the functions of the brain in many different ways and is often used to test multiple aspects of cognitive performance. Tests such as the Flanker Task efficiently measure more than one brain function like executive functioning and attention, while others, like most of the popular reasoning tests (abstract, inductive, deductive), just measure one. The tests that target a single facet of cognitive performance might give more accurate, specific data on that one facet but other tests could be more useful if a general understanding of the many functions of the brain is desired. Our study calls for a test that measures cognitive performances that may be affected by physical exertion, such as attention and memory.

Measuring physical exertion has become much easier in recent years as the availability of wearable technology has increased. Some devices give general data such as heart rate or steps while others provide detailed information on physical exertion by including attachments such as the pulse oximeter. In particular, our study requires a test that can accurately measure the acute changes in physical exertion throughout one session of exercise.

3 Methodology

To study the effects of exercise on cognitive abilities, we selected physical education classes for participants in our exercise group. For the control group, participants were also selected based on which students could get class credit for participation. We also selected cognitive tests that would be relevant for testing attention and memory for our participants. Fitness devices were also chosen for accurate measurements of participants' heart rate and movement data. Survey questions were created to determine participants' demographics, and finally an experimental methodology was created.

3.1 Participants

A total of 62 students participated in this experiment (33 male, 28 female, 1 not reported), between the ages of 17-26 from Worcester Polytechnic Institute. Participants in the exercise condition (30) were recruited from the school's Physical Conditioning and Plyometrics classes. One Physical Conditioning class and five Plyometrics classes participated in the study. Students were given credit for that session of their gym class. Participants in the resting condition (32) were recruited from the university's Psychology and Interactive Media & Game Development programs. These students were given class credit for their participation. All participants provided informed consent before the experiment began.

3.2 Study Design

This study design was a 1x1, with one independent variable with one level. The independent variable was the absence or presence of exercise in the form of a gym class. Students were monitored using an activity tracker to assess how much energy and effort was exerted during the workout. The control condition watched a television show of their choice for the same duration of time. The dependent variable in this design was the change in cognitive function in the specific aspects of attention and memory, before and after exercising. This was operationalized through administering two short cognitive tests before and immediately after the participants either exercise or watch a show. The study also included a brief survey at the end to examine different explanations for cognitive changes.

3.3 Exercise Types

Participants that chose to exercise for the experiment were already enrolled in a university gym class that ran the course of the user study. These gym classes met twice a week for 50 minutes each.

- Plyometrics: This course teaches the use of body weight to develop personal strength and conditioning.
- Physical Conditioning: This course teaches basic strength training principles and techniques. Students develop and implement an individualized conditioning program.

3.4 Cognition Tests

We selected two cognitive tests based on attention and memory to capture two different kinds of cognitive abilities. We selected the Corsi test to measure memory and the Flanker test to measure attention. Both of these tests were taken from the website PsyToolKit (Stoet, 2017).

Memory

To test cognitive function, a variation of the Corsi Test was used to measure memory. To perform this test, participants were shown a variety of purple boxes on the screen. Once the user started the test, the program highlighted one box in yellow which would only be shown for about half a second. Participants had to pay attention and remember which boxes were highlighted and in what order. They were then asked to replicate what was shown by clicking the boxes in the same order and pressing a green box at the bottom of the page to signify that they are done. The Corsi test ended when either they got 9 boxes correct or failed a trial twice in a row. Due to random assignments, about half of the participants took the Corsi Test first and others took the Flanker test first.

To score this task, data of their Corsi Span (how many boxes they could select in the correct order) was recorded by the program.

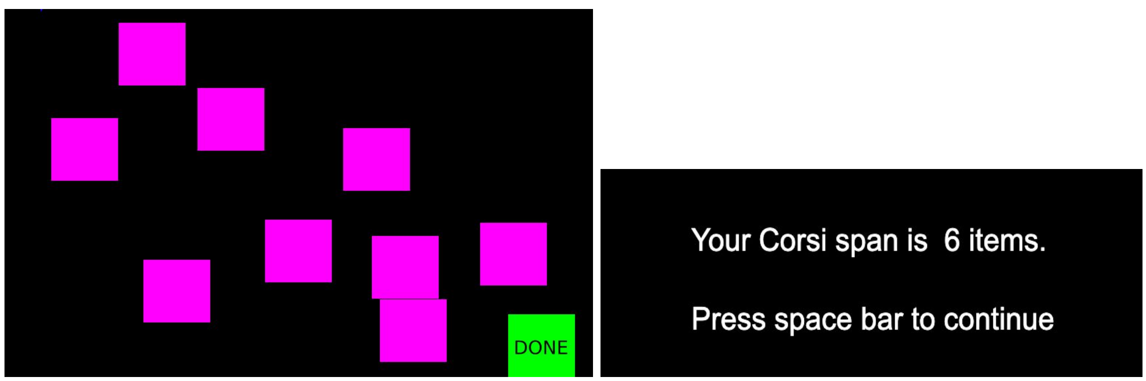

Figure 1: Corsi Test Screenshots

Figure 1 above is a screenshot of the Corsi test, where participants have to memorize which purple boxes were selected. The Corsi test's results are also shown at the end.

Attention

To further test cognitive function, the Flanker test was performed with a variation of the Flanker task to measure attention. To perform this test, participants were shown a row of 5 letters on the screen. They were asked to differentiate whether the middle letter was an "X" or "C" by pressing the "A" key on the keyboard or whether the middle letter was a "V" or "B" by pressing the "L" key on the keyboard. The middle letter was displayed in a row of other letters being similar or different to the middle letter, showing congruent and non-congruent stimuli, respectively. The participants were given approximately 50-60 trials. To score this test, data of how many correct answers and reaction time in milliseconds (time it took between the trial appearing and participants pressing any key) for all trials was recorded by the program. Any reaction time above 2 seconds is considered incorrect and the test will move on.

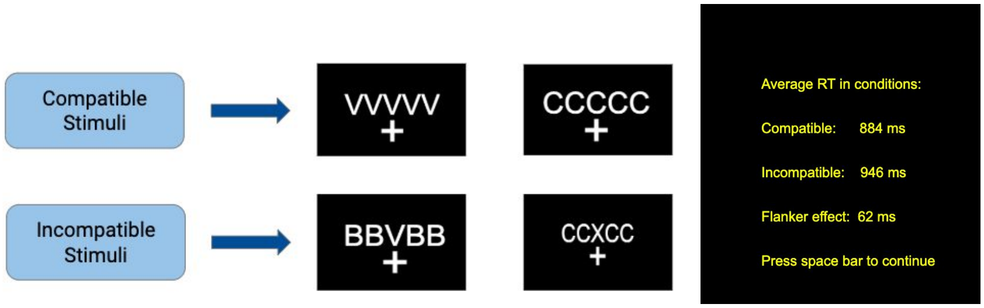

Figure 2: Flanker Test Screenshots

Figure 2 (left) shows screenshots from the Flanker tests, with compatible and incompatible stimuli for the test-taker to react to. The test results are also shown (right) which display the reaction time for compatible and incompatible stimuli as well as the Flanker effect (difference between incompatible and compatible stimuli).

3.5 Activity Tracking

Activity trackers were used to measure the level of physical exertion reached by the participants. A FitBit Inspire HR was used to measure the participant's heart rate during their activity, and the Axivity AX3 was used to record their movement. After the exercises, data from each device was extracted and put onto a computer for analysis, and each trial was separated based on a unique number assigned to the participant since no other identifying information was acquired.

3.6 Experimental Procedures

The test was preceded by an informed consent letter that the willing participants signed. Then, each participant was provided with a laptop to perform their initial cognitive testing. After providing consent, each participant completed both of their cognitive tests and their results were saved online for the investigators to later examine. After both tests were completed, the participants put the exercise devices on their wrists and proceeded to their normal gym workout. Immediately following the workouts, each participant returned to the computers and completed the same two cognitive tests. Once the tests had been saved, everyone was provided with a debriefing document and the first session of the experiment was completed. For the control group, the experiment consisted of the same survey and debriefing. The control group participants also completed the same cognitive tests, but instead of exercise, partook in a leisure activity of watching a show on a computer for 30 minutes while sitting down, and then took the tests again. The survey at the end was used to gather basic information about the participant. The table below shows the questions they were asked as well as the possible responses.

Table 1: Participant Survey Questions

Question	**Response**
How many hours do you sleep a night on average?	1 to 10
How many times do you exercise per week on average?	0 to 7
What kind of exercise do you perform? Check all that apply.	Multiple choice (can select multiple): Running, walking, yoga, sprinting, weight lifting, swimming, elliptical, stairmaster
Do you play a varsity or club sport?	Yes or No
How fit are you?	1 (not very fit) to 5 (very fit)

How many hours ago did you last eat?	1 to 8
Do you have asthma?	Yes or No
Have you taken any over the counter or prescription drugs in the past 24 hours? If yes, please list.	Free response
On average, how many cups of coffee/tea/energy drinks do you drink in a week?	1 to 7
How many cups of coffee/tea/energy drinks have you had today?	0 to 5
Do you drink?	Yes or No
Do you smoke?	Yes or No
How old are you?	Free response
Gender identity	Male, Female, or Other

Questions about exercise were mainly used to test our second hypothesis, which says that people who exercise regularly will show more improvement to cognitive function than people who do not. We also asked other questions we thought might have a correlation with their scores on the cognitive tests, including how much sleep they had, how much caffeine they had, and if they had taken any medications.

3.7 Data Analysis

The data from these survey questions was collected and analyzed with the rest of the data to see if there are any trends with the answers and their exercise data or cognitive data.

Data from the exercise devices was downloaded to the investigators' computers for statistical analysis. The results from the participant survey were used to analyze how the weekly activity of the participants may have affected their cognitive testing and physical exertion during the exercise sessions. It was also used to anonymously compare the participants' performances. The exercise device data was used in correlation with the cognitive testing data, which was also downloaded to the investigators’ computers. This analysis examined whether or not the participants' cognitive performance improved based on the amount of exertion during exercise. The experimental analysis also discovered whether or not the control group or the exercise group improved the more from the initial tests to the post-activity/inactivity tests.

4 Results

In this chapter, we will be reporting the results from the experiment. First, we compared the exercise group and the control group's population demographics in order to examine differences that may affect their cognitive performance and/or activity data. Then, we reported on the cognitive test results within groups and between groups. Next, we examined if there were relationships between participants' habits and cognitive performance. Finally, we looked at both groups' overall activity based on heart rate and accelerometry and finally compared these numbers to their cognitive performance.

4.1 Sample Demographics

A total of 62 students participated in this experiment (33 male, 28 female, 1 not reported), between the ages of 17-26 at WPI. We first categorized the survey data to see the characteristics of our two samples and see if we should consider other variables that may account for changes in cognitive abilities.

Table 1: Survey Data Attributes

Question	Exercise Group: Mean (St Dev)	Control Group: Mean (St Dev)
Sleep, (hours)	6.87 (1.04)	6.95 (1.15)
Exercise, (times per week)	4.27 (1.17)	3.68 (1.83)
Fitness Level, (0-5)	3.3 (0.79)	2.88 (1.01)
Last Meal Time, (hours)	3.76 (2.25)	3 (1.93)
Caffeine, (cups per week)	2.9 (2.36)	4 (2.27)
Today's Caffeine, (cups)	1.22 (0.42)	1.43 (0.68)
Age, years	19.83 (1.73)	19.69 (1.12)

In Table 1, the exercise group and control group both recorded very similar average hours of sleep per night of about 7. The exercise group reported more frequent exercise and reported a higher self-evaluation of fitness than the control group. The highest difference in average is the number of times participants drink caffeinated drinks per week, with the control group drinking about 1 more caffeinated beverage per week than the exercise group on average. Both groups on average reported eating about 3-4 hours before the experiment and reported drinking about 1 to 1.5 cups of caffeinated drinks the day of the experiment. Both groups are about 19-20 years old on average. There is no statistically significant difference in the groups between hours of sleep (p=0.2699), weekly exercise (0.13899), reported fitness level (0.2071), last meal time (0.2298), age (0.06098) or weekly caffeine intake (0.0666). There is a significant difference between caffeine intake on the day of exercise (0.03426).

Table 2: Survey Data - Habits

Question	**Exercise**		**Control**	
	Yes	**No**	**Yes**	**No**
Do you play a varsity or club sport?	4	26	9	23
Do you have asthma?	4	26	4	28
Have you taken any over the counter or prescription drugs in the past 24 hours? If yes, please list.	5	25	10	22
Do you drink?	17	13	16	16
Do you smoke?	5	25	1	31

In Table 2, there are slightly more people in the control group that play a varsity sport and 4 people have asthma in each group. Double the amount of people in the control group recorded taking prescription drugs in the past 24 hours. The majority of drugs recorded in both groups include hormonal medicine (e.g. birth control, levothyroxine, spironolactone) and psychiatric medicine (e.g. stimulants, antidepressants, sleep aids, and bipolar medications). About the same amount of people drink in both groups with an even split within the control group and close to that in the exercise group. Four more people smoke in the exercise group than in the control group.

Table 3: Survey Data - Exercise

Question: What kind of exercise do you perform? Check all that apply.	**Exercise**	**Control**
Running	20	16
Walking	13	17
Yoga	3	3
Sprinting	9	2
Weightlifting	21	15
Swimming	0	1
Elliptical	7	6
Stairmaster	2	3

In Table 3, the top three types of exercise performed by the most participants in both groups are running, walking and weightlifting. The largest difference between both groups is 7 more participants in the exercise group recording that they exercise by sprinting than in the control group.

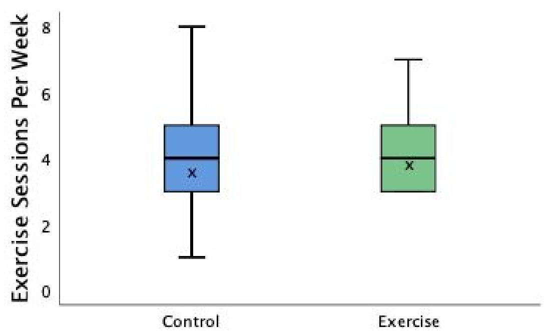

Figure 1, Comparison of how many times participants exercise on average per week.

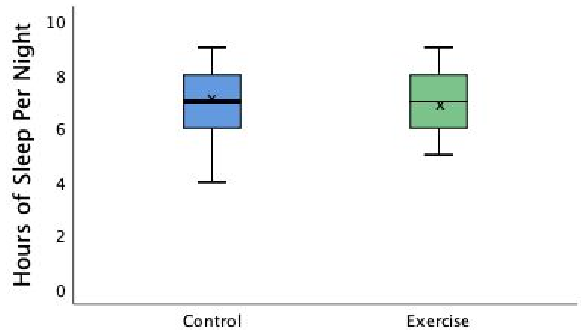

Figure 2, Comparison of how many hours of sleep participants get a night on average.

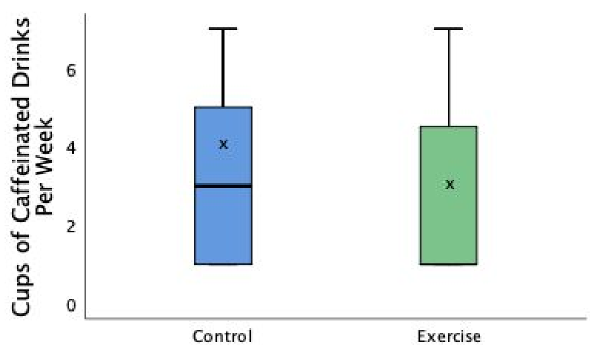

Figure 3, Comparison of how many cups of caffeinated drinks participants drink on average per week. self reported fitness on a scale from 1-7.

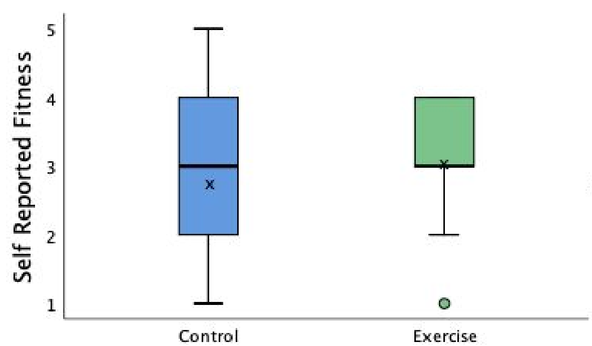

Figure 4, Comparison of the participants self reported fitness on a scale from 1-7.

Figures 1-4 show boxplots of the key survey attributes. For all of the boxplots, the box represents the inter quartile range comprised of the lower quartile (25%), or the bottom line on the box, the middle line represents the median, the "x" represents the mean, and the top line represents the upper quartile (75%). The error bars represent the extent of the range, showing maximum and minimum values.

Figure 1 and Figure 2 demonstrate that control and exercise groups both have a median of 4 exercise sessions per week and about 7 hours of sleep per night with a similar mean for both.

Figure 3 shows the control group drinks about 1 more cup of a caffeinated drink per week than the exercise group. Figure 4 shows that when participants were asked how fit they are on a scale of 1-7, the control group had a wider range of responses compared to the exercise group, with a lower mean than the exercise group but overall similar medians. These boxplots demonstrate that both the exercise and control group had no extreme differences in average amount of exercise per week and hours of sleep per night but showed slight differences in caffeine intake and self reported fitness.

4.2 Cognitive Test Results

Corsi

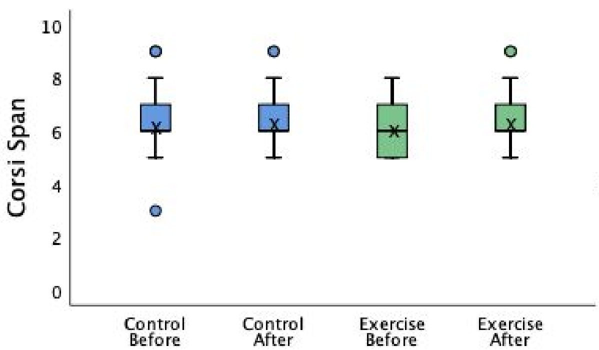

Figure 5, Comparison of the corsi span in a corsi test using raw data from both conditions.

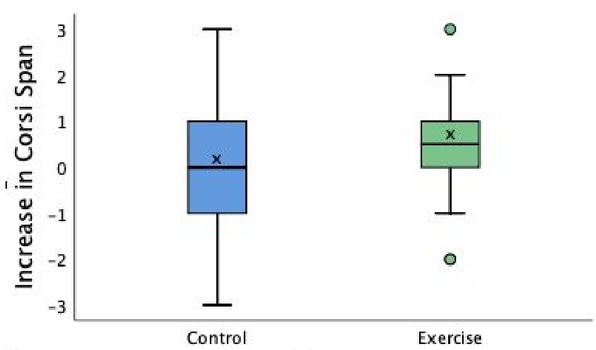

Figure 6, Comparison of the increase in corsi span (corsi span after - corsi span before) in a corsi test.

Based on mean, median and interquartile range (IQR), Figure 5 shows the corsi span for the control group before and control group after visually is about the same and the exercise group seems to have slightly smaller IQR and higher mean. There was a significant difference within groups before and after Corsi spans (higher corsi spans indicate increased memory). In Figure 6, even though this difference is not significant, the exercise group has a higher mean and median increase in Corsi span compared to control.

Four two-sample t-tests assuming equal variance and one paired two samples for means t-test were performed to determine the relationships between Corsi test results in the control and exercise groups. The following values *M* are for the mean value and *SD* for the standard deviation.

Table 4: Corsi Test Results

	M (SD) (units: Corsi span)	**t**	**p**
Test 1: exercise before vs after	6.133(0.937); 6.633(1.066)	-2.186	0.0185
Test 2: control before vs after	6.375(1.2378); 6.656(1.0351)	-1.0865	0.1428
Test 3: exercise after vs control after	6.6333(1.066); 6.656(1.0351)	-0.0858	0.4659
Test 4: exercise before vs control after	6.1333(0.9371); 6.656(1.0351)	-2.0806	0.0209
Test 5: Corsi span improvement, exercise vs control	0.5(1.2526); 0.2813(1.4643)	0.6301	0.26551

Table 4 shows t-test results from both the Corsi and Flanker tests. The tests are labeled by number and referenced below and the first group is before the semicolon while the second group is after.

The group referred to as "first" is always the first group of data reported in the sentence and "second" is the group after. The units for the tests are the Corsi span, which is how many boxes the user could identify in a row without failure. For the exercise group Corsi span results before vs the exercise group after (Test 1 from Table 4), there was a significant difference between the first group (M=6.133; SD=0.937) and the second group (M=6.633; SD=1.066); t(-2.186), p=0.0185. For the control group before vs the control group after (Test 2 from Table

4), there was not a significant difference between the first group (M=6.375; SD=1.2378) and the second group (M=6.656; SD=1.0351); t(-1.0865), p=0.1428. For the exercise group after vs the control group after (Test 3 from Table 4), there was not a significant difference between the first group (M=6.6333; SD=1.066) and the second group (M=6.656; SD=1.0351); t(-0.0858), p=0.4659. For the exercise group before vs the control group after (Test 4 from Table 4), there was a significant difference between the first group (M=6.1333; SD=0.9371) and the second group (M=6.656; SD=1.0351); t(-2.0806), p=0.0209. For the difference in Corsi span after minus before (exercise vs control) (Test 5 from Table 4), there was not a significant difference between the first group (M=0.5; SD=1.2526) and the second group (M=0.2813; SD=1.4643); t(0.6301), p=0.26551

Flanker

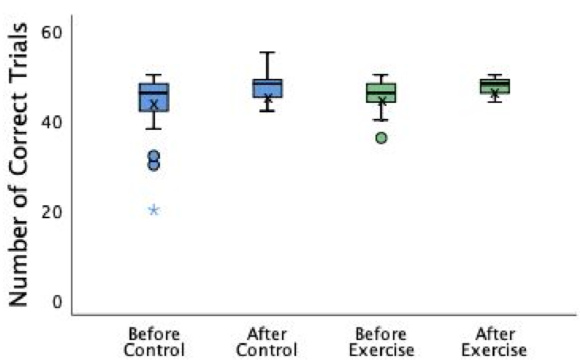

Figure 7, Comparison of the number of correct trials in the flanker test using raw data from both conditions.

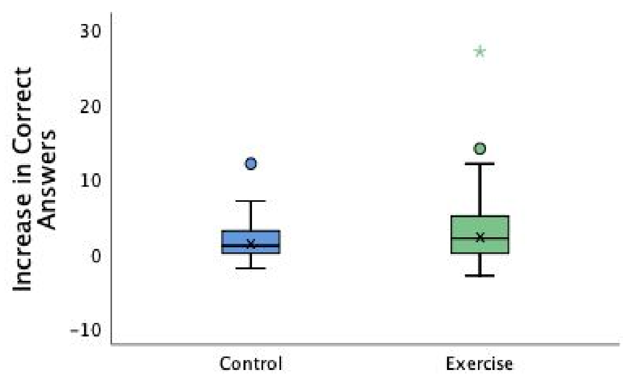

Figure 8, Comparison of average response time for correct responses in a flanker test using raw data from both conditions.

In Figure 7, shows significant results, within conditions, for the number of correct trials in the Flanker test (higher trend meaning more correct responses). Both control and exercise groups slightly improved in their mean and median from before to after. Figure 8 shows no visual difference between groups for their increase in correct answers before and after, and this is proven as the t-test showed no significant difference between groups.

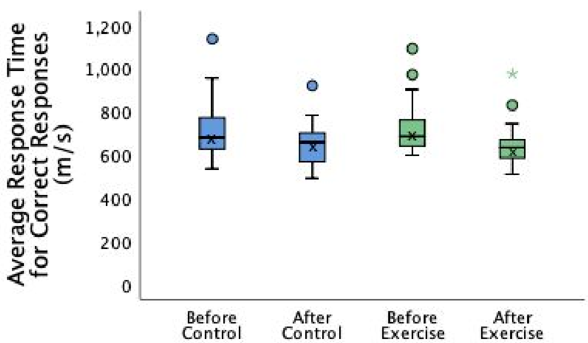

Figure 9, Comparison of the increase in correct responses (correct responses after - correct responses before) in a flanker test.

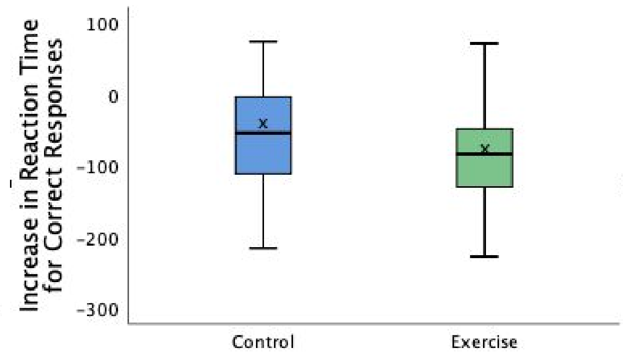

Figure 10, Comparison of the increase in response time (rt) for correct responses (rt for correct after - rt for correct before) in a flanker test.

Figure 9 shows that the average response time for correct responses (lower trend meaning faster response time) within groups got smaller, meaning a faster reaction time before and after the exercise or control condition. This is shown in the paired t-tests for the control and exercise groups, with a significant difference of before and after scores in both. In Figure 10, although not deemed significantly different by a t-test, the data shows a larger visual decrease in reaction time for the exercise group compared to control in both the mean and median.

Table 5: Flanker Test Results

	M (SD) (units: Correctness- Flanker answers correct. Response time- ms)	**t**	**p**
Test 6: exercise correct before vs after	43.7(6.5923); 47.3333(2.6695)	-3.4217	0.00094
Test 7: exercise response before vs after	724.1333(112.102); 637.8(98.5622)	6.8457	<.001
Test 8: control correct before vs after	45.7(3.16391); 47.4667(1.8889)	-3.32665	<.001
Test 9: control response time before vs after	714.2(128.4608); 650.8(90.3123)	4.51697	<.001
Test 10: difference in correctness, exercise vs control	3.6333(5.8161); 1.7667(2.90877)	1.5723	0.06067
Test 11: difference in response time, exercise vs control	-86.333(69.075); -63.4(76.878)	-1.2154	0.11457

Four two-sample t-tests assuming equal variance and two paired two samples for means t-tests were performed to determine the relationships between Flanker test results in the control and exercise groups. For the exercise number of correct answers before vs exercise correct answers after (Test 6 from Table 4), there was a significant difference between the first group (M=43.7; SD=6.5923) and the second group (M=47.3333; SD=2.6695); t(-3.4217), p=0.00094. For the exercise response time (in milliseconds) before vs the exercise response time after (Test 7 from Table 4), there was a significant difference between the first group (M=724.1333;

SD=112.102) and the second group (M=637.8; SD=98.5622); t(6.8457), p<.001. For the control correct answers before vs the control correct answers after (Test 8 from Table 4), there was a significant difference between the first group (M=45.7; SD=3.16391) and the second group (M=47.4667; SD=1.8889); t(-3.32665), p<.001. For the control response time for correct answers before vs the control response time for correct answers after (Test 9 from Table 4), there was a significant difference between the first group (M=714.2; SD=128.4608) and the second group (M=650.8; SD=90.3123); t(4.51697), p<.001. For the difference in correctness after minus before (exercise vs control) (Test 10 from Table 4), there was not a significant difference between the first group (M=3.6333; SD=5.8161) and the second group (M=1.7667; SD=2.90877); t(1.5723), p=0.06067. For the difference in response time after minus before (exercise vs control) (Test 11 from Table 4), there was not a significant difference between the first group (M=-86.333; SD=69.075) and the second group (M=-63.4; SD=76.878); t(-1.2154), p=0.11457.

4.3 Habits vs Cognitive Performance

Pearson correlation tests were run to determine the relationships between the amount of caffeine the participants consumed the day of the experiment and their cognitive testing results. Values “p" are the Pearson correlation score and values in parenthesis after the correlation score are the p values of the test. All tests that measure response time have a negative value if the participant improved because this indicates a faster response time while taking the tests.

Table 6: Habits vs Cognitive Performance

	r (Corsi; Flanker correct; Flanker response time for correct)	P (Corsi; Flanker correct; Flanker response time for correct)
Caffeine	-0.13; -0.07; 0.29	.32; .56; .02
Sleep	0.16; -0.36; 0.22	.22; .005; .09
Exercise	0.15; 0.01; -0.23	.25; .94; .08

Table 6 shows the r and p values for caffeine, sleep and exercise habits vs cognitive results.

The correlation for caffeine vs Corsi span score after the experimental session minus before the experimental session is r=-0.13 (p=.32). The correlation for caffeine vs Flanker correct answers after the experimental session minus before the experimental session is r=-0.07 (p=.56). The correlation for caffeine vs Flanker response time for correct answers after the experimental session minus before the experimental session is r=0.29 (p=.02).

Pearson correlation tests were run to determine the relationships between the hours of sleep the participants got the night before the experiment and their cognitive testing results. The correlation for hours of sleep vs Corsi span score after the experimental session minus before the experimental session is r=0.16 (p=.22). The correlation for hours of sleep vs Flanker correct answers after the experimental session minus before the experimental session is r=-0.36 (p=.005). The correlation for hours of sleep vs Flanker response time for correct answers after the experimental session minus before the experimental session is r=0.22 (p=.09).

Pearson correlation tests were run to determine the relationships between the number of times participants exercise on a weekly basis and their cognitive testing results. The correlation for number of times of exercise per week and Corsi span score after the experimental session minus before the experimental session is r=0.15 (p=.25). The correlation for the number of times of exercise per week and the difference between Flanker correct answers before and after the experimental session is r=0.01 (p=.94). The correlation for number of times of exercise per week and the difference in Flanker response time before and after the experimental session is r=-0.23 (p=.08).

4.4 Activity Tracker Results

For each group's activity tracker data (Fitbit, Axivity), metrics important to our hypotheses were gathered. These include mean, median, and standard deviation. However, unlike previous tests, each group's data was not analyzed together but instead each participant was analyzed individually. Every trial gathered data from each activity tracker, and since everybody has different biology and workouts, we analyzed them individually. Instead, a cumulative distribution function (CDF) graph showed each participant's data while preserving the experimental grouping.

A CDF is defined by the following formula: $F_X(x) = P(X \leq x)$. The right hand side of the equation is the probability that a real-valued random variable X will take a value less than or equal to x. Thus, a CDF can be used to effectively group data and see trends. The user's heart rate percentage reached was calculated as a percent of their maximum heart rate, based on the American Heart Association's (AHA) recommendation of the age subtracted from 220 (AHA, 2015). In the case of Figure 11 below, the vertical axis represents the fraction of users' highest heart rate percentage reached less than or equal to the value on the horizontal axis.

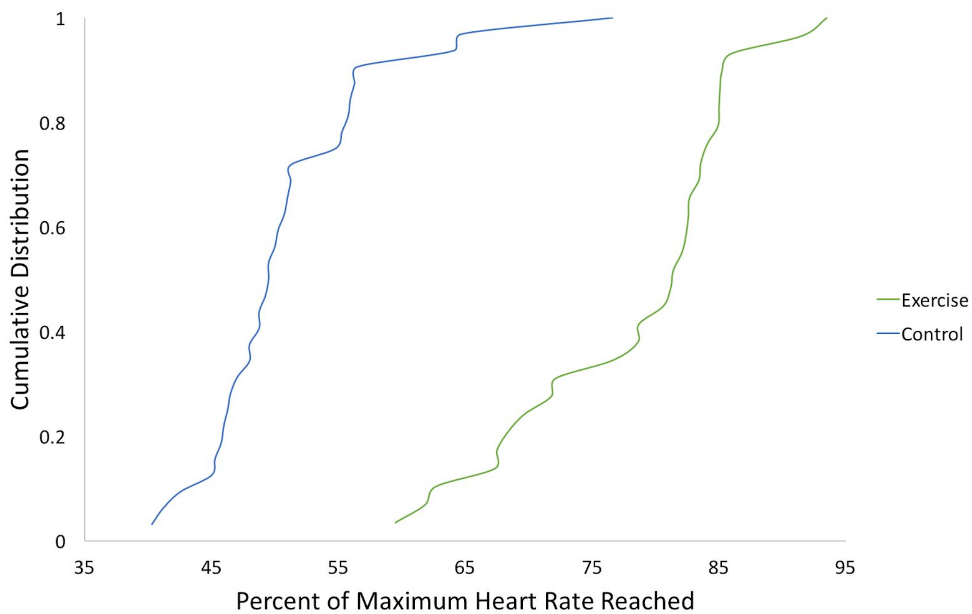

Figure 11, CDF of each user's percent maximum heart rate reached in the control group (blue) and exercise group (green).

CDF graphs were also made for the Axivity results. For each Axivity device, it measures the triaxial acceleration every 10 ms (100 Hz). For the analysis, each recording was combined into a three dimensional vector and its magnitude summed to find the mean. The mean is their average acceleration through their exercise or control session.

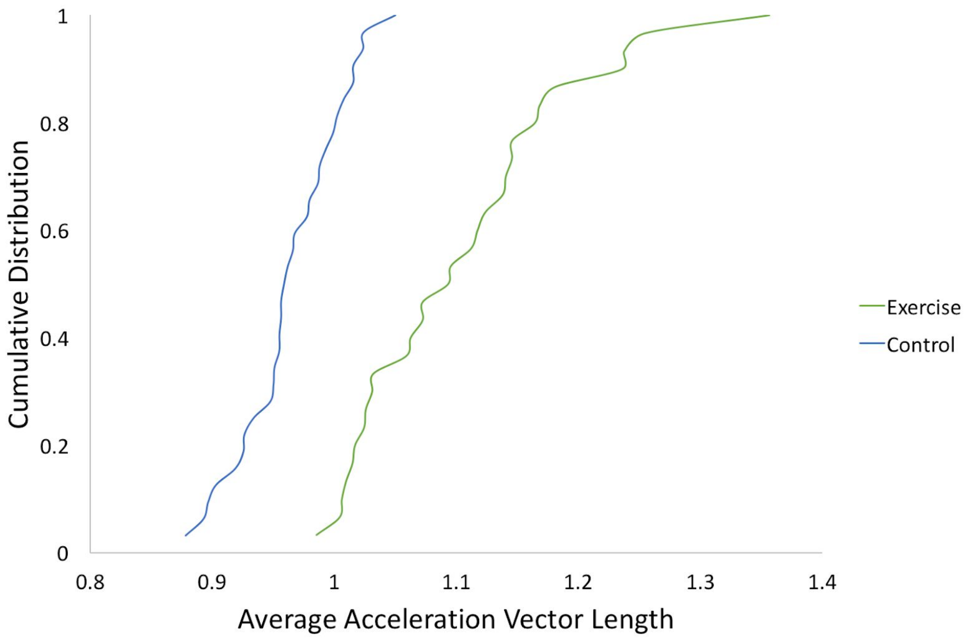

Figure 12, CDF of each user’s average acceleration vector length in the control group (blue) and exercise group (green).

Figure 12 shows the fraction of a user’s acceleration magnitude that is less than or equal to the values on the horizontal axis. Note that Axivity records acceleration in units of g, with $g \approx 9.81\ m/s^2$. Any numbers on the horizontal axis must be multiplied by g to get the acceleration in m/s^2 .

The participants’ average heart rate was graphed alongside the average acceleration to see if there is a correlation between acceleration and heart rate. Graphing both of these on the same graph, with acceleration on the horizontal axis and heart rate on the vertical axis, there is a slight upward trend. This trend is stronger in the exercise group than the control group, but both show a correlation between acceleration and heart rate. The correlation coefficient (r) for the exercise group (figure 13) is 0.51, showing a moderate positive correlation, while the correlation coefficient for the control group (figure 14) is 0.11, showing a much smaller positive correlation.

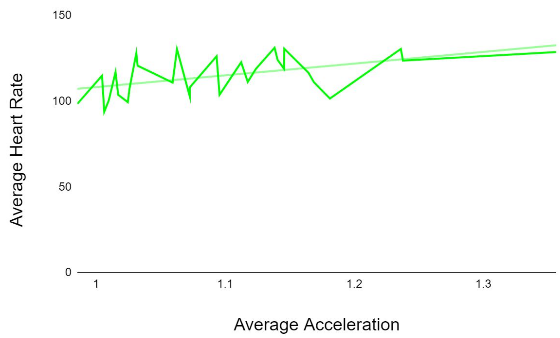

Figure 13 plots the average acceleration and average heart rate of the exercise group, along with a line of best fit to see if there are any trends between them ($r = 0.51$).

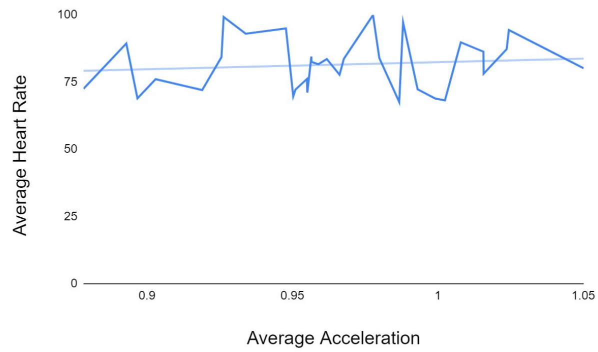

Figure 14 plots the average acceleration and average heart rate of the control group, along with a line of best fit to see if there are any trends between them ($r = 0.11$).

4.5 Heart Rate vs Cognitive Performance

Multiple Pearson correlation tests were run in order to examine how a change in heart rate might relate to a change in cognitive performance.

Table 7: Heart Rate vs Cognitive Performance

	r **(Corsi; Flanker correct; Flanker response time for correct)**	**p**
% of max HR achieved	0.04; 0.37; -0.19	0.76; 0.004; 0.15
Mean HR	-0.05; 0.35; -0.26	0.97; 0.006; 0.04
Median HR	0.04; 0.17; -0.15	0.76; 0.19; 0.25
% of session above 70% of HR	-0.06; 0.46; -0.09	0.65; <0.001; 0.49

Table 7 shows the *r* and *p* values for these tests. The first value is the correlation with increase in Corsi span. The second value is for the increase in correctness in the Flanker test. The third is for the increase in response time for correct answers in the Flanker test.

Pearson correlation tests were run to determine the relationships between the percentage of maximum heart rate that participants achieved during activity and their improvements on the cognitive tests. This was done to observe how participants that may reach a point of exhaustion did on their tests. The correlation between percentage of maximum heart rate achieved and the difference of Corsi span before and after activity is r=0.04 (p=0.76). The correlation between percentage of maximum heart rate achieved and the increase in correct answers in the Flanker test after activity minus before activity is r=0.37 (p=0.004). The correlation between percentage of maximum heart rate achieved and the increase in response time for correct answers in the Flanker test after activity minus before activity is r=-0.19 (p=0.15).

Pearson correlation tests were run to determine the relationship between the mean heart rate that participants achieved during activity and their improvement on the cognitive tests. The correlation between mean heart rate achieved and the increase in Corsi span after activity minus before activity is r=-0.05 (p=0.97). The correlation between mean heart rate achieved and the increase in correct answers in the Flanker test after activity minus before activity is r=0.35 (p=0.006). The correlation between mean heart rate achieved and the increase in response time for correct answers in the Flanker test after activity minus before activity is r=-0.26 (p=0.04).

Pearson correlation tests were run to determine the relationships between the median heart rate that participants achieved during activity and their improvements on the cognitive tests. The correlation between median heart rate achieved and the increase in Corsi span after activity minus before activity is r=0.04 (p=0.76). The correlation between median heart rate achieved and the increase in correct answers in the Flanker test after activity minus before activity is r=0.17 (p=0.19). The correlation between median heart rate achieved and the increase in response time for correct answers in the Flanker test after activity minus before activity is $r=-0.15$ ($p=0.25$).

The Fitbit data was also analyzed to see how much (as a percent) of the participant's activity session took place while they were at or above 70% of their maximum heart rate. This was done because the American Heart Association defines vigorous exercise as at least 70% of the maximum heart rate (AHA, 2015), and we wanted to observe whether or not vigorous exercise caused an increase or decrease in cognitive performance. The correlation between the percent of activity in which the participants were over 70% of their max heart rate and the increase in Corsi span after activity minus before activity is $r=-0.06$ ($p=0.65$). The correlation between the percent of activity in which the participants were over 70% of their max heart rate and the increase in correct answers in the Flanker test after activity minus before activity is $r=0.46$ ($p<0.001$). The correlation between the percent of activity in which the participants were over 70% of their max heart rate and the increase in response time for correct answers in the Flanker test after activity minus before activity is $r=-0.09$ ($p=0.49$).

4.6 Axivity Data vs Cognitive Performance

Pearson correlation tests were run to determine the relationships between the amount of movement that the participants achieved during activity and their cognitive performance during the tests.

Table 8: Axivity Data vs Cognitive Performance

	r (Corsi; Flanker correct; Flanker response time for correct)	**p**
Accelerometry data	0.09; 0.19; -0.31	0.49; 0.15; 0.016

Table 8 shows the summary correlation data between accelerometry data and cognitive testing results.

Pearson correlation tests were run to determine the relationships between the amount of movement that the participants achieved during activity and their cognitive performance during the tests. The correlation between the average acceleration vector length and the increase in Corsi span after activity minus before activity is r=0.09 (p=0.49). The correlation between average acceleration vector length and the increase in correct answers in the Flanker test after activity minus before activity is r=0.19 (p=0.15). The correlation between average acceleration vector length and the increase in response time for correct answers in the Flanker test after activity minus before activity is r=-0.31 (p=0.016).

5 Conclusion

The goal of this study was to examine whether exercise can show cognitive benefits in college-aged students. Previous studies have been done on this topic (Chang et. al., 2012), but none have been done on the college student age group. These studies have typically been done on older populations, but this study would like to show college students that there is a benefit to exercise. The Centers for Disease Control and Prevention showed that there is a decline in the amount of regular exercise from high school and college students (Centers for Disease Control and Prevention, 2016). The main hypothesis is that acute physical activity will improve cognitive performance more than a period of inactivity, and a second hypothesis is that people who exercise more regularly will experience greater improvements in cognitive performance after exercise than people who exercise less regularly.

The first hypothesis was tested by comparing the results of the cognitive tests of the exercise group and the control group. Significance tests show that there was no significance between the two groups before and after ($p \geq 0.05$), however the graphs visually show a difference. The Corsi span before and after for each group shows no statistically significant difference (p=0.27), but visually the box and whiskers plot for exercise shows a higher mean, median, and 1st quartile than the control group. Similarly, the graphs show a visual improvement in response time for the exercise group compared to the control group. Even though there was no significant difference between groups, the exercise group before and after was significant (p=0.02), while the control group before and after was not significant (p=0.14), which shows that exercise causes a significant difference in results compared to not exercising.

An addition to the first hypothesis is that people who are working to exhaustion will do worse on their cognitive tests due to fatigue. This was tested by analyzing the heart rate data and calculating what percentage the user is above 70% of their max heart rate, which is the value defined by the American Heart Association for "vigorous exercise". The correlation between this and difference in Corsi span is -0.06, showing only a slight decrease in results. The correlation between this exhaustion percent and difference in Flanker (test of acuity) correct answers is 0.46 ($p < 0.001$), showing a moderate correlation, meaning users actually did better the longer they were above 70%. The correlation between exhaustion and Flanker average response time before and after is -0.09, meaning they answered only slightly faster. The correlations between accelerometry data and the three cognitive testing results (Corsi, Flanker correct, Flanker response time for correct) were 0.09, 0.19, -0.31 respectively, showing no statistically significant correlations.

The second hypothesis was tested by running Pearson correlation tests on the cognitive data and the survey data for amount of exercise per week. The correlation of the amount of exercise per week and Corsi span before and after is 0.15 (p=0.25), showing a very small positive correlation. The correlation between amount of exercise and Flanker correct answers before and after is 0.01, showing almost no correlation. The correlation between amount of exercise per week and Flanker response time for correct answers before and after is -0.23 (p=0.08), showing a small negative correlation (negative is better since the response time is quicker).

Other survey data was also correlated with the cognitive test results. The correlation between amount of caffeine consumed on the day of the experiment and Corsi span before and after the experiment was -0.13, showing a small negative correlation. The correlation between caffeine consumed and Flanker correct answers is -0.07, showing a very small negative correlation. The correlation between caffeine consumed and Flanker response time is 0.29, showing a small positive correlation. All these values support that cognitive performance before and after exercise decreases in relation to caffeine increase.

By the *p-value* in the significance tests, the first hypothesis, which says that acute exercise will improve short term cognitive performance more than a period of inactivity, is supported. The second part of the first hypothesis is not supported by the evidence, and it seems like the more vigorous someone works out the better they do, further proving the first hypothesis. Additionally, the data does not support the second hypothesis, which says that people who exercise more regularly will have better cognitive performance than someone who doesn't exercise.

6 Discussion

One major strength of our procedure is that external validity is generally higher than a study completely run in a lab. With this study being done on students participating in their real gym class, the validity of the findings is more likely to extend to situations in the real world. Another reliable component of this study is the Flanker test, as this form of assessing attention is widely used in past studies examining links between exercise and cognition. Our study used two ways of measuring cognitive capabilities and two ways of tracking exercise which also supports the reliability of the results.

Overall, the limitations of this study are mostly related to the limitations of the activity trackers and cognitive tests we chose and some slight limitations due to our procedure. One limitation of this study is when participants showed up even a couple of minutes late to their class it made the procedure rushed. For this reason, we were not able to calculate an accurate resting heart rate, as once the participants put on their Fitbit they were running to their class and not resting. To try to make up for this, we averaged their heart rate over the first two minutes, but even then their results are likely higher than their actual resting heart rate.

Another limitation involves our choice of cognitive tests. The Corsi test is not widely used and has not been validated by past studies. We choose this because of the accessibility of being able to run the test on our laptops and save the data. The site we retrieved it from, PsyToolKit, claimed that the Corsi test is quite similar to the widely used Digit Span test to measure memory. However, the Corsi test is considerably more visual and motor-driven, so participants needed to be able to see quick changes in the color of boxes on the screen and use the mouse to recreate the pattern. It may have hindered participants' ability to do well if they had slow reaction times or mouse skills or if they were simply unable to see the quick color changes.

We also considered the test-retest problem in which participants in both conditions may have improved from pre-tests to post-tests because it is their second time taking the test. Because it is a repeated measure, improvement could be due to practice, known as the practice effect, as participants may have done worse on the pre-tests due to it being the very first time doing it and better on the post-tests due to being exposed to it previously. Although we explained the directions for the test to all participants, the fast-paced nature of the exercise condition may have caused students to rush the pre-test. This could be due to trying to be on time for class and not paying full attention to the tests and beginning before fully comprehending all of the directions. Their post-test may have improved due to it being after class and they had more time to comprehend and may have been more concerned with doing well. This would overall effect the significance of the improvement when comparing conditions. Another explanation for the improvement could be due to the control group still being stimulated and engaged cognitively by the show that they watched.

For the sake of statistical analysis, the findings may have not been significant due to the small sample size. By increasing the sample size, the data would reach a point of saturation and it would be certain whether there was a relationship between exercise and cognitive scores.

7 Future Work

Future work includes priming the participants for cognitive tests. This would aid in removing a “learning curve” for participants’ first trials and allow investigators to know with more certainty whether the participants’ score changes are based on factors other than becoming used to the test. This could be accomplished by allowing the participant’s a practice trial before their actual testing period.

Another area of future work includes measuring exercise by using oxygen saturation levels. This would allow for another measure of physical exertion to help determine the participant’s effort and exhaustion levels. This could be accomplished by adding a pulse oximeter to the participants’ wrists during activity.

An additional variation includes using different cognitive tests. This would allow investigators to examine facets of cognition that the Flanker (largely attention) and the Corsi (largely memory) do not examine. This could be accomplished with online software that offers multiple different kinds of cognitive testing or development of novel cognitive tests that focus on unique aspects of cognition. More specifically, the Corsi test could be replaced with a more widely accepted test of memory.

Another adaptation is performing the same experiment with repeated measures. This would allow investigators to track participants' progress as they adapt to a certain form of exercise. This project would be long-term and more accurately track improvement over time.

One final version of future work includes separating the exercise group based on exercise type. Doing this would allow investigators to examine how different types of exercise may affect a participant’s cognitive performance. This could be accomplished by assigning exercise routines to participants. For example, the following three groups may be interesting to examine: steady-state cardiovascular exercise on a treadmill, high-intensity interval training on a treadmill, and weightlifting.

References

American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription. USA: Lippincott Williams & Wilkins; 2013.

American Heart Association. Know Your Target Heart Rates for Exercise, Losing Weight and Health. (2015). Retrieved from https://www.heart.org/en/healthy-living/fitness/fitness-basics/target-heart-rates

Buckley, Sparks, Papp, Dekhtyar, Martin, Burnham & Rentz. (2017). Computerized Cognitive Testing for Use in Clinical Trials: A Comparison of the NIH Toolbox and Cogstate C3 Batteries. *The journal of prevention of Alzheimer's disease*, *4*(1), 3.

Calestine, J., Bopp, M., Bopp, C. M., & Papalia, Z. (2017). College Student Work Habits are Related to Physical Activity and Fitness. *International Journal of Exercise Science*, 10(7), 1009–1017.

Centers for Disease Control and Prevention. Nutrition, Physical Activity and Obesity: Data, Trends and Maps. 2016.

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, 87-101.

Coffey, Arseneau-Bruneau, Zhang & Zatorre. (2019). The Music-In-Noise Task (MINT): A Tool for Dissecting Complex Auditory Perception. *Frontiers in Neuroscience*, *13*.

Colcombe, S. J., Erickson, K. I., Raz, N., Webb, A. G., Cohen, N. J., McAuley, E., & Kramer, A. F. (2003). Aerobic Fitness Reduces Brain Tissue Loss in Aging Humans. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 58(2), M176-M180.

Committee on Psychological Testing, I. V. (2015). Cognitive Tests and Performance Validity Tests. *NCBI*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK305230/.

Davranche, Hall & McMorris. (2009). Effect of Acute Exercise on Cognitive Control Required During an Eriksen Flanker Task. *Journal of Sport and Exercise Psychology*, *31*(5), 628-639.

Emery, Schein, Hauck & MacIntyre. (1998). Psychological and Cognitive Outcomes of a Randomized Trial of Exercise Among Patients with Chronic Obstructive Pulmonary Disease. *Health Psychology*, *17*(3), 232.

EMotion mobile FAROS sensor for ECG, HRV. High Resolution, Small and Lightweight. (n.d.). Retrieved September 19, 2019, from http://ecg.biomation.com/faros.htm

Eriksen & Eriksen. (1974). Effects of Noise Letters Upon Identification of a Target Letter in a Non-search Task. *Perception and Psychophysics* 16: 143–149.

Farnsworth. (2016). N-back to Basics: Learning and Memory with the N-back Test. *Imotions*. Retrieved from https://imotions.com/blog/n-back-test/.

Hogan, C. L., Mata, J., & Carstensen, L. L. (2013). Exercise Holds Immediate Benefits for Affect and Cognition in Younger and Older adults. *Psychology and Aging*, 28(2), 587-594. doi:10.1037/a0032634

Hogan, M., Kiefer, M., Kubesch, S., Collins, P., Kilmartin, L., & Brosnan, M. (2013). The Interactive Effects of Physical Fitness and Acute Aerobic Exercise on Electrophysiological Coherence and Cognitive Performance in Adolescents. *Experimental Brain Research*, 229(1), 85-96.

IPhone X - Technical Specifications. (n.d.). Retrieved September 19, 2019, from https://support.apple.com/kb/sp770?locale=en_US

Kamijo, K., Hayashi, Y., Sakai, T., Yahiro, T., Tanaka, K., & Nishihira, Y. (2009). Acute Effects of Aerobic Exercise on Cognitive Function in Older Adults. *The Journals of Gerontology: Series B*, *64B*(3), 356–363. https://doi.org/10.1093/geronb/gbp030

Khan, A. M. (2012, February 24). Recognizing Physical Activities Using the Axivity Device. *ResearchGate.* Retrieved from https://www.researchgate.net/publication/256404940_Recognizing_Physical_Training_Exercises_Using_the_Axivity_Device

Kramer, A. F., Hahn, S., Cohen, N. J., Banich, M. T., McAuley, E., Harrison, C. R., … Colcombe, A. (1999). *Ageing, Fitness and Neurocognitive Function*. Retrieved from https://www.nature.com/articles/22682#rightslink

Mead & Drasgow. (1993). Equivalence of Computerized and Paper-and-pencil Cognitive Ability Tests: A Meta-analysis. *Psychological Bulletin, 114*(3), 449-458. Retrieved from http://dx.doi.org/10.1037/0033-2909.114.3.449

Olympic Medical Institute Validates Polar RS800 Running Computer And Training System | Polar USA. (2014, February 25). Retrieved September 19, 2019, from https://web.archive.org/web/20140225052618/http://www.polar.com/us-en/about_polar/news/polar_RS800

Penedo, Frank J, and Jason R Dahn (2005, March). “Exercise and Well-Being: a Review of Mental and Physical Health Benefits Associated with Physical Activity.” Current

Opinion in Psychiatry, *U.S. National Library of Medicine*, Retrieved from www.ncbi.nlm.nih.gov/pubmed/16639173.

President Bush Discusses National President's Challenge. (2008). Retrieved September 19, 2019, from https://georgewbush-whitehouse.archives.gov/news/releases/2008/03/20080320-9.html

PsyToolkit - A software package for programming psychological experiments using Linux.

PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. Stoet, G. (2017). *Teaching of Psychology, 44(1)*, 24-31.

Samsung Gear S2—Full phone specifications. (n.d.). Retrieved September 19, 2019, from https://www.gsmarena.com/samsung_gear_s2-7678.php

Scarpina & Tagini. (2017). The Stroop Color and Word Test. *Frontiers in psychology*, *8*, 557.

Shanhong, Liu. (2019, May 22). Topic: Fitness & Activity Tracker. Retrieved September 19, 2019, from Www.statista.com website: https://www.statista.com/topics/4393/fitness-and-activity-tracker/

Specifications | Samsung Galaxy S10e, S10 & S10+. (n.d.). Retrieved September 19, 2019, from The Official Samsung Galaxy Site website: https://www.samsung.com/global/galaxy/galaxy-s10/specs/

Sternberg & Sternberg. (2016). Cognitive Psychology. *Nelson Education*.

Stoet, G. (2010). *Behavior Research Methods, 42(4)*, 1096-1104.

Tomporowski, Phillip D. (2002, December 4). Effects of Acute Bouts of Exercise on Cognition. Retrieved from https://www.sciencedirect.com/science/article/pii/S0001691802001348

Wahid A, Manek N, Nichols M, Kelly P, Foster C, Webster P, Kaur A, Friedemann Smith C, Wilkins E, Rayner M, et al. Quantifying the Association Between Physical Activity and Cardiovascular Disease and Diabetes: A Systematic Review and Meta-Analysis. J Am Heart Assoc. 2016;5:pii: e002495.

Willetts, M., Hollowell, S., Aslett, L., Holmes, C., & Doherty, A. (2018). Statistical Machine Learning of Sleep and Physical Activity Phenotypes from Sensor Data in 96,220 UK Biobank Participants. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-26174-1

123 Test. (2019). Abstract Reasoning Test. 123 Test. Retrieved from https://www.123test.com/abstract-reasoning-test/.