# BUILDING ADOLESCENT EXECUTIVE FUNCTIONING THROUGH A PLAY-BASED CURRICULUM

**JADE S. L. OLDFIELD**

Graduate Certificate in Educational Neuroscience, University of Calgary, 2022

A thesis submitted

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

NEUROSCIENCE

Department of Neuroscience

University of Lethbridge

LETHBRIDGE, ALBERTA, CANADA

BUILDING ADOLESCENT EXECUTIVE FUNCTIONING THROUGH A

PLAY-BASED CURRICULUM

JADE S. L. OLDFIELD

| | | |
|---|---|---|
| Date of Defence: | December 20, 2024 | |
| Dr. Robbin Gibb<br>Thesis Supervisor | Professor | Ph.D. |
| Dr. Claudia Gonzalez<br>Thesis Examination Committee Member | Professor | Ph.D. |
| Dr. Lance Grigg<br>Thesis Examination Committee Member | Associate Professor | Ph.D. |
| Dr. David Euston<br>Chair, Thesis Examination Committee | Professor | Ph.D. |

## DEDICATION

To everyone who supported me in this undertaking; from Kayla Marcotte who suffered the onslaught of my never ending questions with patience and grace, to the undergraduate students who worked beside me throughout the process (namely, Kaileb, Rylee, Jonathan and Ameet), to my amazing professors in all departments who were forever supportive. Most of all, however, to my forever understanding husband who carried much on his shoulders to help me get to this point. I could name people endlessly, but the truth is that everyone involved in this journey has had a profound impact on not only the outcome of this degree, but the trajectory of the next.

## ABSTRACT

Executive functions (EFs) are critical skills that impact all aspects of a persons' life and are more predictive of life success than academic achievement (Alloway & Alloway, 2010). While there are different perspectives on how to categorize EF, herein we focus on the three distinct, yet interrelated behaviors of inhibition (the ability to withhold a response), working memory (the ability to keep in mind and manipulate information; Diamond, 2013), and cognitive flexibility (the ability to move from one frame of reference to another; Best & Miller, 2010). Of particular interest is the development of EF during adolescence as this is currently a large gap in the research. Adolescence is a dynamic and fundamental period of growth and change for the brain, during which EF matures (Gee & Casey, 2015). If an environment is not optimal for EF maturation, individuals may have poorer life outcomes (Tooley et al., 2021; Zysset et al., 2018). Therefore, the current study outlines a detailed and accessible play-based intervention that shows promise in helping adolescents build EF and hypothesizes that participants will see improvements in inhibition, working memory, and cognitive flexibility. While significant results are seen in the post-test results, they appear to be limited to the domain of inhibition or inhibitory control.

Keywords: executive function, adolescent, play, inhibition

**ETHICS STATEMENT**

Work described in this thesis received research ethics approval from the University of Alberta Research Ethics Board, Project Name "BUILDING BRAINS TOGETHER," No. Pro00120933, November 28, 2022. See Appendix A.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ADHD | Attention Deficit Hyperactivity Disorder |
| BBF | Building Brains and Futures |
| BBT | Building Brains Together |
| BRI | Behavioural Regulation Index |
| BRIEF-SR | Behavioural Rating Inventory of Executive Function© - Self-Report |
| CON | Congruent |
| CF | Cognitive Flexibility |
| DCCS | Dimensional Change Card Sort |
| EF | Executive Functions |
| ER | Emotional Regulation |
| fMRI | Functional Magnetic Resonance Imaging |
| GEC | General Executive Composite |
| GLM | General Linear Model |
| IC | Inhibitory Control |
| INC | Incongruent |
| IQR | Interquartile Range |
| LBBT | Lego® Brick Building Task |
| MC | Motor Control |
| MI | Metacognition Index |
| ms | Milliseconds |
| NS | Not Significant |
| PA | Physical Activity |
| PFC | Prefrontal Cortex |
| PVT | Picture Vocabulary Test |
| RM | Repeated Measures |
| RT | Response Time |
| SPSS | Statistical Package for Social Sciences |
| SR | Self-Report |
| STD | Standard |
| T1 | Time One (Pre-Testing) |
| T2 | Time Two (Post-Testing) |
| TOT | Total |
| WM | Working Memory |

# CHAPTER 1: INTRODUCTION

Executive functions (EFs) refer to a range of functions that impact a person's ability to interact with their environment. There are three distinct, yet interrelated behaviors: inhibition, the ability to withhold a response, working memory, the ability to hold and manipulate information in your mind, and cognitive flexibility, the ability to move from one frame of reference to another (Best & Miller, 2010). As reviewed in Gee & Casey's work on adolescents, adolescence is a dynamic and fundamental period of growth and change for the brain, during which EF matures (Gee & Casey, 2015). If an environment is not optimal for EF maturation, individuals may have poorer life outcomes (Tooley et al., 2021; Zysset et al., 2018) and may even come into contact with the justice system. My research asks whether there are play-based approaches to aid the development of EF during adolescence. As this has previously been demonstrated with preschool children (Coelho et al., 2020; Gibb et al., 2021), this study proposes to extend and assess the impact of play-based curricula on EF by analyzing the impact of this intervention on a group of Grade 7 students during a critical period in their neurobiological development: adolescence.

## Theory and Hypothesis

*Adolescent cortical development: A critical period*. Adolescence is characterized by multifaceted changes. Crews et al. propose that adolescence is a critical period associated with cortical maturation of the prefrontal cortex (PFC; Crews et al., 2007). As a result of these developments, there is an increased opportunity to leverage social and educational environments, related to the development of EF. We propose that one way to leverage this critical developmental period is to use play as a vehicle of change.

Play is a fundamental and evolutionarily advantageous aspect of life that helps children to build skills, schemas, and provides youth with an opportunity to exercise executive function skills within a safe environment (Yogman et al., 2018). The question this thesis seeks to answer is if a selection of games, played a minimum of three times a week in the classroom, over the course of 6 weeks, will improve EF. It is hypothesized that the purposeful inclusion of games, chosen to build executive functions, will enhance the EF skills of inhibition, working memory, and cognitive flexibility (Sala & Gobet, 2016). These changes will be measured through five psychometric assessments (the Stroop, Dimensional Card Sort, Picture Vocabulary Test, Flanker Inhibitory test and the Word List Sort Working Memory task), the Behavioural Rating Inventory of Executive Function, Self-Report (BRIEF-SR) survey, and a Lego® brick building motor task (LBBT).

## Background

### *Critical Periods*

Critical periods or windows, as referred to in this review, refer to a broader concept of sensitive periods in cerebral plasticity originally stemming from the experiments in the 1960-70s conducted by Hubel and Wiesel. Their Nobel Prize winning research demonstrated that a lack of visual input during a specific developmental window in young brains of kittens caused changes in visual cortex networks (Hubel & Wiesel, 1963). They concluded that visual experiences during this critical period are essential for normal neural circuit development and that normal visual experiences after this period cannot correct the abnormal circuits (Hubel & Wiesel, 1970).

Sensitive periods are defined as the window(s) of time where the effect of certain experience(s) on the development of the brain is extraordinarily impactful and can affect and modulate neural circuits. During sensitive periods, neural circuits are highly adaptable and possess multiple genetically encoded potential connection patterns (Ismail et al., 2017). They can commit to a specific pattern based on stimuli, remaining open to change if the preferred stimulus is absent and later reintroduced, though this adaptability diminishes with age. Thus, while all critical periods are sensitive periods, not all sensitive periods are critical. It is important to recognize that critical and sensitive periods are not absolute and vary in their timing and characteristics across different neural subtypes, networks, regions, and even behavioral phenotypes (Ismail et al., 2017).

Adolescence, which is defined as the developmental stage that begins with the onset of puberty and ends when individuals reach adulthood (Jaworska & MacQueen, 2015), also has sensitive periods for memory, stress, mental health, and effects of drug use (Fuhrmann et al., 2015). Adolescence is a developmental stage that has an increase in brain plasticity, pubertal maturation, and behavioral shifts, such as increased independence and a stronger focus on peer relationships (Blakemore & Mills, 2014). It is also a time when the onset of psychopathology is most common (Kessler et al., 2005), particularly among youth who have experienced stressful life events (McLaughlin et al., 2012).

### ***Executive Functions***

**Inhibition**. Inhibitory control enables individuals to regulate their attention, actions, thoughts, and emotions, thereby modulating the ability to control internal drives, resist external temptations, or—especially in the case of adolescents—to resist impulsive behaviours (Diamond, 2013). The fundamental EFs of inhibition, working memory and cognitive flexibility, allow us to make deliberate choices and modify our behavior rather than aimlessly following our habitual responses (Diamond, 2013). Inhibition encompasses multiple components, such as response inhibition, self-control, the ability to resist impulsive responses, interference control, selective attention, and cognitive inhibition (Diamond, 2013).

Inhibitory control of attention, also known as interference control at the perceptual level, allows us to focus selectively on what we choose and disregard other stimuli. Despite this skill, prominent and salient stimuli, such as loud noises or visual movement, can naturally capture an individual's attention. This is driven by the stimuli's properties and is commonly referred to as exogenous, bottom-up, automatic, stimulus-driven, or involuntary attention (Diamond, 2013). Conversely, we can deliberately ignore specific stimuli and focus on others based on our goals or intentions, known as selective or focused attention, attentional control, attentional inhibition, endogenous, top-down, active, goal-driven, voluntary, volitional, or executive attention (Diamond, 2013).

Self-control, an essential aspect of inhibitory control, allows individuals to manage behavior and emotions to resist temptations and impulsive actions. Self-control allows individuals to avoid indulging inappropriately, avoid excess, and to adhere to social norms. Additionally, self-control involves maintaining focus on tasks despite distractions or the allure of immediate gratification, often linked with delaying gratification for greater future rewards, which is essential for completing long-term goals (Diamond, 2013). In laboratory settings, young children tend to respond hastily, but teaching them to wait can enhance their performance across various tasks requiring inhibitory control. The subthalamic nucleus is crucial for curbing impulsive responses (Diamond, 2013).

Inhibitory control presents a significant challenge for young children, with their performance reflecting greater difficulty in inhibiting dominant responses rather than holding multiple associations in mind. This difficulty persists throughout childhood but gradually improves during adolescence. Early-life inhibitory control is a strong predictor of lifelong outcomes. Longitudinal studies have shown that children with better inhibitory control are more likely to exhibit positive behaviors as teenagers and adults, such as staying in school, making fewer risky choices, and maintaining better physical and mental health (Diamond, 2013). However, inhibitory control tends to decline with normal aging. Older adults demonstrate deficits in inhibiting visual and auditory distractions and show less suppression of irrelevant information than younger adults. These inhibitory-control challenges in aging are evident across various tasks, indicating a decline in inhibitory abilities over time (Diamond, 2013).

**Working Memory.** Working memory (WM) refers to the ability to retain and manipulate information within the mind, even when it is no longer directly perceivable. It is a fundamental aspect of executive function and is pivotal in various cognitive tasks (Diamond, 2013). Verbal and nonverbal (visual-spatial) are the two main types of WM, each serving distinct functions in our cognitive processes.

Verbal working memory is essential for comprehending language, whether it is a single sentence or a complex paragraph. It facilitates mental arithmetic, task organization, incorporating new information into existing frameworks, and considering alternative solutions. Moreover, verbal WM underlines our capacity for reasoning, enabling us to draw connections between concepts and fostering creativity by facilitating the recombination of ideas (Diamond, 2013). Nonverbal WM, on the other hand, deals with spatial and temporal information, allowing us to retain and manipulate visual-spatial data; it is crucial for tasks requiring the manipulation of mental imagery, such as mentally rearranging objects or envisioning spatial relationships (Diamond, 2013). Additionally, as per Diamond (2013), nonverbal WM enables us to integrate conceptual knowledge into decision-making processes, drawing from our memory of past experiences and future aspirations.

Distinguishing WM from short-term memory is vital, as they serve distinct roles in cognitive processing. While short-term memory involves the passive retention of information, WM requires active manipulation and processing of that information. These forms of memory are associated with different neural pathways, with WM heavily relying on the dorsolateral prefrontal cortex (Diamond, 2013). Developmentally, short-term memory tends to mature earlier and more rapidly than WM, as evidenced by performance discrepancies in tasks requiring memory retention versus manipulation across different age groups (Diamond, 2013).

**Cognitive Flexibility.** Cognitive flexibility (CF) refers to the ability to adapt swiftly and flexibly to different situations, viewpoints, and frames of reference. This involves shifting perspectives spatially or interpersonally, inhibiting previous viewpoints, and activating new ones. Moreover, CF entails adjusting to altered demands or priorities, acknowledging errors, and seizing unexpected opportunities. There exists a considerable overlap between cognitive flexibility, creativity, task-switching, and set-shifting. CF is characterized by adaptability rather than rigidity (Diamond, 2013).

When being assessed for cognitive flexibility using various psychological tests, young children often struggle with switching between mental sets and exhibiting attentional inertia–getting stuck in focusing their attention on a previous way of thinking about a stimulus. Adults, on the other hand, typically demonstrate greater ease in maintaining consistent responses within a task block. However, switching between mental sets remains more challenging than consistently inhibiting dominant responses. Overall, cognitive flexibility demands overcoming inertial tendencies to navigate between different perspectives or ways of thinking effectively (Diamond, 2013).

### ***Play and Executive Functions***

Play can improve EF skills through two pathways, indirect and direct. Directly, play may improve executive functions by providing youth and children with cognitive challenges and stimulation (Diamond & Ling, 2016). When engaging in pretend play, children exercise all aspects of executive function. They must remember the storyline and their role, pause their current actions to take on their character, and be receptive to suggestions from their peers or changes in the story. Therefore, pretend play allows children to practice different components of executive function (Blair & Diamond, 2008).

Indirectly, play tends to be enjoyable, which is linked to a decrease in stress, anxiety, and loneliness (Diamond & Lee, 2011). Because stress has a negative impact on the prefrontal cortex (Arnsten, 2009) and children's EF development (Blair et al., 2011), it seems that play is crucial for those indirect positive effects. It is clear that play and the positive development of executive functions are linked. The purpose of this literature review is to support the creation of a curriculum of games that promotes the positive development of executive functions.

### ***Physical versus Table-Top Games***

To categorize the various games reviewed in this paper, the two groups of games are characterized as tabletop (non-physical) and physical activity games. To provide a distinct understanding of terms used in this paper, 'physical activity' is defined as "any bodily movement produced by skeletal muscles that results in energy expenditure" (Caspersen et al., 1985, p. 126). While exercise and physical activity (PA) are terms often used interchangeably, they are distinctly different and should be clarified. Exercise is a "subset of physical activity that is planned, structured, and repetitive and has as a final or an intermediate objective the improvement or maintenance of physical fitness" (Caspersen et al., 1985, p.128). This distinction is important as it relates to the direct and indirect benefits of play. For many, exercise–such as running–may be seen as a chore that is done to maintain ones' health; on the other hand, a recreational game of soccer–which also encompasses running–could be seen to be more fun and playful than just going for a run, thereby inducing the indirect benefits of play. Adults with more fully developed prefrontal cortices have the ability to utilize the top-down EF skills explained above to rationalize, justify and convince themselves of the importance and benefit of the run, helping to maximize the health benefits. However, as adolescents' EF skills are still developing, it can be more challenging to convince them of the innate benefits of exercise. With this in mind, the experimental curriculum includes games that have a physical activity component to help leverage the benefits of both PA and the proposed positive EF effects.

**Conclusion**

The concept of critical periods is not new, and while it is widely accepted that the preschool years are some of the most important developmentally, adolescence as a critical period is still gaining traction (Gee & Casey, 2015; Kolb et al., 2023). Continued research, such as this study, is needed to further explore the plasticity that occurs during puberty, and how these neurobiological changes impact EF. Furthermore, adolescence is a period of critical social development, and while play may look slightly different during this time, teenagers still benefit from the social, emotional and neurological aspects developed through play. Consequently, the importance of play cannot be overlooked with respect to adolescent development and the development of EF.

# CHAPTER 2: LITERATURE REVIEW

## Introduction

Executive functions are critical skills with broad implications; research shows EFs have correlations to mental health and learning disorders, are predictive of educational attainment, and are even predictive of IQ (Alloway & Alloway, 2010). Given the importance of these skills, there is a need, and a desire, for interventions that can provide remedial support. This chapter provides the overview for the "Building Brains Together" (BBT) adolescent game curriculum, which was designed to help improve EF through play. BBT, originally founded as Building Brains and Futures (BBF), began as a research program that designed and tested a play-based curriculum to improve EF in preschool children (see Gibb et al., 2021). The BBF program was founded by Gibb, Harker and Gonzalez at the University of Lethbridge and when tested showed significant positive results in measures of children's EF. Building on this program's success, a second program was proposed that targets the second critical window of development: adolescence (Coelho et al., 2020).

## Literature Review

### *Play*

Research on the interaction of play-behavior and executive functions is limited, both in childhood and adolescence, therefore, this literature review focuses on existing research that considers both play and executive functions in humans whenever available. Not surprisingly, much research and defining has focused on what constitutes animal play. Burghardt (2011) collates many definitions to create five criteria of play:

1. Play, in form and function, has an evolutionary advantage, despite not contributing immediately to survival.
2. The act of play is spontaneous, voluntary and done for the sake of play alone.
3. The act of play may mirror functional behaviours, but is different in that it may be exaggerated, incomplete or is otherwise modified.
4. Behaviours are repetitive, so as to be similar or identifiable, but not so rigid as to be prescriptive.
5. Play occurs when the subject is in a safe space, with adequate food, clothes and is not stressed.

Similar to Burghardt, Eberle (2014) presents similar criteria for the definition of play, albeit slightly more simple and human-centric:

1. Play exists simply for play's sake.
2. Players play because they want to.
3. Play exists in a separate place, although those places can be many.
4. Play must be fun.
5. All types of play have rules.

Eberle continues to explore the vastness and ethereal nature of play, including Burghardt's definitions. The difficulty in this literature review, as also explored with respect to EF and social development, is that there is limited research with respect to play in adolescence. Furthermore, much of the literature that does exist focuses on the spontaneous and unstructured aspects of play; these criteria become seemingly less prevalent in adolescence as social structure and EF skills begin to develop. In the current study, play occurs in the classroom, albeit being organized by the teacher. Other typical adolescent play occurs online but is not spontaneous in that a group of friends agrees to log in at a certain time. In addition to this, teenagers have the ability to choose how they want to play; they can choose solo play (ie. video game campaign modes where they engage with the content alone), collaborative play (ie. working together with others towards a shared goal), and competitive play (ie. playing with at least one other person with the intention of reaching an end goal before they do). While the literature is still developing in this area, there are studies that show distinct differences between the aforementioned modalities, suggesting that social play increases positive mood and the enjoyment of the activity (Kaye & Bryce, 2014), whereas solo play can involve less distractions and can increase engagement with the content (Ruipérez-Valiente & Kim, 2020). Adolescence is a developmental period that still requires much research and conceptual reflection, and play is among the concepts that will need reconsideration. For the purposes of this thesis, we will focus on competitive and collaborative play involving two or more youth. Accordingly, we will define adolescent play as any activity in which two or more adolescents willingly engage, that (1) is either self-directed or group-directed, but can also be adult-initiated, (2) has rules and structure, but (3) its primary purpose is connecting with others over a shared goal (whether collaborative or competitive), and (4) while the activity may have underlying educational, developmental or social benefits, "play exists for its own sake" (Eberle, 2014, p. 215) and invokes joy.

Play has been shown to be essential across species (Yogman et al., 2018; Pellis & Pellis, 2009). In humans, play serves multiple functions, such as the development of social skills, language abilities and theory of mind (Lillard, 2015). Despite its importance, free-play (as defined by unscheduled, unstructured play [Lillard, 2015]) is becoming less available to children, both in the school-day and in their daily lives (Ginsburg et al., 2007).

Play has been a topic of interest for years. Pellis & Pellis (2007; 2009; et al., 2010) have studied the form and function of play in animals quite extensively, despite early skepticism of its value, but what about humans? Gray (2009) proposes that play is evolutionarily advantageous, and that it encouraged equality amongst early hunter-gatherer tribes. Furthermore, Piaget and Vygotsky advocated for play as a fundamental part of development (Piaget, 1951; Vygotsky, 1978).

While play has been well studied in younger children, and many theories and classifications abound, less study has been conducted on adolescent play. When searching the Web of Science, using topic keywords of “play,” “teenager,” and “adolescent(s),” the search returned only 482 results, most of which had little-to-no connection to the concept of play, with the exception of video game playing. Therefore, the literature in this review also encompasses the author’s extensive search for research about play that extends beyond childrens’ first years of school.

***Adolescent Development***

The adolescent brain is quite possibly one of the most complex biological objects in the known universe. Despite advances in neuroscience and psychology, the critical developmental period of the adolescent brain still holds much mystery. In 1999, Giedd et al. conducted a thorough and complex longitudinal study, in which MRI scans of participants were taken at two-year intervals between the ages of 4-21. While this paper highlighted many notable findings, in sum, it was able to establish adolescence as a truly critical developmental period.

Much like play, the definition of adolescence is ethereal, escaping the grasp of those who seek to pin it down. While some older definitions posit that the defining feature of adolescence centers around the transition to independence, Sawyer et al. (2018) apply a broader definition of adolescence. They suggest that due to both biological and social determinants, this developmental period spans between 10- and 24-years old. For the purposes of this review, we will adopt Sawyer et al.'s (2018) view and definition, asserting that puberty starts earlier, and that adolescence in fact lasts longer than in the past.

**Puberty & Hormones.** The onset of puberty has often been seen to coincide with the modern definition of adolescence. However, as is discussed in Gluckman (2006), the initial age of puberty has been shifting younger for both males and females, likely due to the availability and amount of nutritious food in the environment. This finding is also in line with Sawyer et al.'s position that puberty starts around 10 years of age (2018). The importance of these findings is that the gonadotropin hormones that are released during puberty have an effect outside of secondary sex characteristics: they also contribute to brain plasticity. As suggested by Giedd et al. (1999), this second increase in gray matter, which occurs roughly one year earlier in females–in line with pubertal development–could "herald a critical stage of development when the environment or activities of the teenager may guide selective synapse elimination during adolescence" (p. 863). Thus, while our understanding of the adolescent brain is still quite limited, there is significant potential in leveraging this second developmentally critical period.

**Brain Development**. In addition to the importance of the onset of hormones, brain development during adolescence also affords many opportunities. The area of the brain undergoing the most significant construction during adolescence is the frontal lobe, with the occipital and temporal lobes being more fully developed in the teenage years (Giedd et al., 1999). This area of the brain will remain a work-in-progress until at least the mid-twenties (Giedd et al., 1999), however, with the brain awash in hormones, adolescence remains a key time for brain development.

As has been shown in previous MRI studies, there is a general trend in white and gray matter development during the adolescent years: gray matter volume generally declines and white matter increases, with females seeing less increases in white matter during this time (Foulkes & Blakemore, 2018; Giedd et al., 1999; Perrin et al., 2009; Tamnes et al., 2017). Gray matter declines suggest the pruning of cell bodies as the brain eliminates unnecessary connections in an attempt to create efficient connections and increased cortical surface area through the increase in white matter (Kolb et al., 2023; Tames et al., 2017). The tumultuous changes in adolescent brains come with both advantages and disadvantages: the brain is primed for plasticity and learning. However, adolescent brains are also primed for social interactions and risk taking. This duality has been referred to as the imbalance model (Ernst et al., 2006) as well as the dual-systems model (Somerville et al., 2010; Steinberg, 2010).

**Social Development.** Another important facet to adolescent development is the change that children and youth undergo with respect to social development as they enter the teenage years. While science has not yet teased apart the impact of the changes created through the influx of hormones and the neural development in the PFC, Blakemore and Choudhury (2006) suggest that a two-way interaction exists. The authors suggest that brain development affects adolescent social cognition, but the development of social cognition–and what is perceived as socially salient–also then impacts further brain development. The authors continue, suggesting that perspective taking, an integral part of social cognition and social development, undergoes significant reorganization, similar to the PFC, during adolescence. Furthermore, we know that the development of both EF skills and emotional regulation (ER) is taking place during adolescence; while cognitive EF (working memory, cognitive flexibility and cognitive inhibition) are related to more cognitive success, ER (considered to overlap with inhibition; Diamond, 2013) is essential for healthy social interactions. Interestingly, in a time where social interactions and perceptions are critical to adolescent self-esteem, pubertal teens' ability to accurately identify and match emotions to facial expressions declines compared to their pre- and post-pubertal peers (approximately before age 10-12, and after age 16-17; Blakemore, 2008). This is just one example of how turbulent the adolescent stage truly is. As a result of this biopsychosocial flux that puberty brings, adolescents need modeling and support to help build both EF skills and pro-social interaction capabilities.

***Executive Functions***

Executive functions are a set of cognitive skills with a neurobiological root, predominantly located in the frontal lobes. EF skills have been broadly defined in the past, but the author sides with Lehto et al.'s definition (2003; Miyake et al., 2000; Diamond, 2013) that focuses on the three main domains of: working memory, cognitive flexibility and inhibition. These skills are crucial for life success, and in fact, are predictive of such (Alloway & Alloway, 2010). EF skills are also predictive of math, language and general academic achievement (Diamond & Lee, 2011; Jacob & Parkinson, 2015). With the frontal lobes under development during the adolescent years, and with EF skills being critically important for academic success, this developmental period presents an opportunity whereby interventions can be leveraged to improve EF.

As previously discussed, the adolescent brain undergoes dramatic changes in gray and white matter volume, as well as heterogeneous changes in specific subcortical structures such as the amygdala, cerebellum and hippocampus (Foulkes & Blakemore, 2018), all of which have important implications with respect to learning. With respect to EF, however, the most important neurobiological factor is the relationship between synaptic pruning and the development of the prefrontal cortex (PFC). As shown by Gogtay et al. (2004), the “higher-order association cortices” (p. 8174), and their related functions, are the last areas to develop as shown by gray matter maturation. This delayed frontal lobe maturation is thus related to adolescents’ delayed EF abilities. In line with Selemon (2013), it is posited that adolescence is a sensitive period of brain development during which the PFC is undergoing maturation and synaptic pruning, making way for the development of EF.

There has been much research on EF interventions and their potential application, however, due to the range presented on the literature, it can be challenging to determine if an intervention’s effectiveness is due to its modality (ie. technology based interventions such as CogMed©) or its actual efficacy. However, various EF interventions have been proven to be effective (Diamond, 2013; Diamond & Lee, 2011). Diamond & Lee (2011) outline six categories of interventions that have had proven effects on EF in children ages 4-12, these include: computerized training, hybrid training (computer and non-computer games), aerobic exercise, mindfulness, classroom curricula, and Montessori programs. These findings are incredibly promising, but further work needs to be done to better understand the underlying correlations between these differing intervention modalities.

In general, more work also needs to be done to address the limited data available on adolescent populations and EF. According to Jacob & Parkinson, in their 2015 meta-analytic study, only 8% of studies focused solely on adolescents, while 14% included youth across multiple age categories. While much remains to be understood regarding both EF and the adolescent brain, it is clear that important correlations exist between academic achievement and EF, and that adolescence is a period in which the brain, and in particular the PFC, is undergoing drastic changes; changes which can hopefully be leveraged for the benefit of the adolescent brain.

### *Physical Activity and EF*

Physical activity (PA) has also been correlated with EF skills, however, like much of the research into EF, the population tested is predominantly of preschool or elementary school age. Also, there are often variable definitions of "physical activity." For example, Sember et al. (2020) conducted a meta-analysis on interventions involving childrens' PA. However, they note that many studies consider only the amount and/or frequency of PA, and do not factor in the intensity. Despite the conclusion that PA does offer at least small positive effects on academic performance, they note that intensity, among other factors, does play a role. Having defined exercise in Chapter 1, here we will specify what we refer to as PA. As per Caspersen et al.'s definition (1985), PA involves bodily movement that results in the expenditure of energy, whereas exercise includes the aforementioned conditions, with the addition of being a planned and structured activity with the objective of improving or maintaining fitness (p. 127). While the two concepts are not always separated in the literature, for the purposes of this article, we are largely concerned with the concept of PA and not that of regimented exercise.

**Motor Control & EF.** The relationship between motor control (MC)—the ability to consciously control both fine and gross motor skills—and EF has been debated for some time, however, various lines of research have shown a correlation between children's and adolescent's MC, cognition and EF. Luz et al. (2015) showed that a relationship between MC and EF exists: children with better MC also have improved cognition, but the authors suggest that these relationships may be modulated by processing speed. Albuquerque et al. (2022) showed that MC and EF are more strongly associated in younger children, but an association also exists in older children (9-11 years old). Rigoli et al. (2012) take the correlation a step further and provide evidence that MC and EF may be related through cerebellar functioning. Finally, Gonzalez et. al (2014) looked at the relationship between motor performance and EF and found that preferential right hand use, as well as using the right space of the task, predicted EF abilities. Likewise, EF ability was also a predictor of right hand use. With the connections between MC, cognition and EF becoming more established, it's clear that activities that employ and enhance MC have the potential to improve cognition, and more importantly, EF.

**PA, Cognition, and Academic Achievement.** It has been shown time and time again that physical activity has a multitude of benefits that cross a variety of health dimensions, and cognition is no exception to this. While research on physical activity within adolescent populations is sparse, general research, or research within other age groups, supports the cognitive benefits associated with PA and exercise (Dishman et al., 2006; Shephard, 1997). These benefits include an increased arousal state, the release of brain-derived neurotrophic factor ([BDNF], Li et al., 2017) as well as increased activation of the PFC (Best & Miller, 2010; Hillman et al., 2008). Numerous studies have thus linked improved cognition and academic achievement with increased physical activity (Davis et al., 2007; 2011), and more specifically with increased EF.

**PA & EF.** In reviewing the literature regarding the brain-based benefits of PA, it was often found that terms such as "cognitive functioning," "academic performance," and "academic achievement" were used to represent EF. For example, Sember et al.'s meta-analysis on PA included many studies which measured EF, however, they then represent these findings as "academic performance" (2020). Similarly, Li et al. (2017) investigated the role of exercise with respect to both cognitive function and academic performance; the authors identified six cognitive parameters, of which three were directly related to EF (ie. “executive function,” “memory,” “attention/concentration,” p. 841).

In sum, most of the literature has found varying effects of PA on cognitive functioning and EF. However, one important point that is noted by Sember et al. (2020) is that contrary to the concern that PA may negatively impact academics by taking time away from them, PA in fact plays a positive role with respect to academic performance. In addition to this, original research (ie. non meta-analyses) from de Bruijn et al. (2018), Luz et al., (2015) and Davis et al. (2007; 2011) point towards positive correlations between PA and EF. Therefore, despite the lack of an explicit causal relationship, the literature most definitely supports the importance and cognitive benefits of PA throughout the lifespan.

**Rationale: Table-Top Games**

The importance of play, especially for younger children, is indisputable; however, due to increased academic pressures, play in adolescence is often sidelined; this play aversion is to the detriment of teens. Based on the current understanding of neurobiology and development, it is clear that the PFC, and the skills it supports (ie. EF), are still under construction in the teenage brain. It is also known that the adolescent brain is very much a “social” brain, and play allows adolescents to both express and build these skills in a safe environment. The BBT Adolescent Curriculum leverages a variety of games chosen for their relation to EF skills. Games include both PA and stationary activities, and are designed to be played face-to-face with varying group sizes. For more information on the gameplay specifically, the game cards can be found on the BBT website (https://www.buildingbrains.ca/; Building Brains Together, retrieved November 14, 2024).

### *Spoons*

Spoons is a tabletop card game that is best played in groups of 4-8 players. A deck of cards is sorted to include a set of four cards of the same face value for each player. The objective of the game is to collect a four-of-a-kind set and then to be the first person to grab one of the spoons from the centre of the table. Once a player grabs a spoon, all other players must follow suit. There are n-1 spoons available (ie. five spoons for six players). Therefore, the person who was unable to grab a spoon is eliminated. Players have four cards in their hand and simultaneously discard and slide one card from their hand to the player on their left - the process of acquiring a card, reviewing it, and then discarding a card all happens in roughly three seconds. Gameplay continues until two players go head-to-head for the final spoon.

This game exercises a number of EFs including: working memory (WM), inhibitory control (IC) and emotional regulation (ER). In addition, despite being a tabletop game with limited PA, players do have to exercise precise motor control which is also related to EF skills as previously discussed. Players exercise their WM in this game by keeping multiple pieces of information in mind while making reasoned judgements (Diamond, 2013). As they quickly review the card they're handed, players need to compare and measure their goal (ie. collecting four of the same cards) against their actual cards (ie. only having two-of-a-kind, and being passed a card that either does, or doesn't, help them achieve their goal). At this point, they also have to employ IC as they inhibit other thoughts and distractions, focusing only on their cards and their goal, as well as not prematurely grabbing a spoon (Miyake et al., 2000). Finally, players must use ER to keep calm during the fast-paced gameplay, especially during the end-of-the-round scramble for spoons and player elimination (Poon, 2018; Zelazo & Carlson, 2012).

***Slap***

Slap is another fast-paced tabletop card game where players readily employ IC as well as WM, cognitive flexibility (CF), and ER. Two decks of cards are shuffled together for roughly four-to-eight players and then all the cards are dealt out evenly amongst players, face down. Players are not allowed to look at their cards and take turns quickly flipping a card and placing it in the centre of the circle of players. Players continue in this fashion, continually employing IC as they wait for a double to turn up (ie. the card on the top of the pile is a six, and the next player also turns over a six from their hand). Once a double is revealed, the first person to *slap* the pile in the centre "wins" all the cards in the centre. Play continues and players are eliminated as they run out of cards in their hands.This describes basic gameplay, but rules can be added so that players must slap for other cards as well, such as when a joker is flipped.

One of the main ways this game taxes IC is that if a player slaps the deck when there is no reason to (ie. they "jump the gun"), they must lose two cards from their hand and place them in the centre. Therefore, there is a "tax" for incorrectly identifying an opportunity to act (Miyake et al., 2000). In addition, players need to balance WM with IC as they must hold some of the recently played cards in mind, while choosing the appropriate response ([go, no-go]; Diamond, 2013). This is directly related to CF as they need to be flexible in their approach according to the card flipped. Finally, like Spoons, this game requires a good deal of ER as players need to control their emotions throughout the rounds of the game (Poon, 2018; Zelazo & Carlson, 2012).

***Concentration (Number & Animal Versions)***

Concentration is a representation game where players are assigned either a number (or an animal, based on the version played), and must "pass" the gameplay around the group of up to 10 players, and it actively employs WM, IC, CF and ER. Play is passed by a player first indicating their number twice (ie. saying “two, two”) or acting out their animal “signal” (ie. flapping of arms to represent “bird”), and then by saying another player’s number twice or acting out their animal signal. Play continues until a player makes a mistake (ie. attempts to pass play to a player that has already been eliminated) and then the player either loses a “life” or is eliminated from the game. The game continues until only one player is left.

This game requires a great deal of working memory; with up to 10 players, players need to remember their own number (or animal), and up to nine more. This knowledge needs to be actively held in mind and accessed rapidly when it’s the player’s turn (Diamond, 2013). At that same time, players need to use CF to shift plans, and update WM, as players are eliminated. This game also requires that players maintain a beat while performing the other EF related tasks. This adds a layer of challenge and supports EF. As reviewed in the motor control section, MC is correlated with EF and other cognitive abilities, and better MC suggests better cognition (Luz et al., 2015). In addition to this, players are not allowed to talk and need to keep a beat going between turns, therefore they need to use IC to block distractions and focus on the task at hand, while inhibiting incorrect responses (Miyake et al., 2000). Lastly, like the aforementioned games, players need to use ER when things don’t go their way.

***Oh Heck***

Oh Heck is a slightly more complicated card game for ideally a group of 5-6 people, that requires players to anticipate outcomes and plan ahead accordingly. Therefore, this game requires a great deal of WM and IC, as well as CF and ER. To play this game, the dealer deals out the entire deck of cards to the players and each subsequent round they deal one less card (ie. for six players, each player would start with eight cards, the next round would start with seven and so on). A “trump” card is placed face up and play continues clockwise from the dealer and before play starts, players need to “bet” on how many tricks (rounds) they will win. The key strategy to this game however, is correctly guessing how many tricks they will win. While winning a trick will earn players one point, it is correctly guessing how many tricks they will win that earns them ten points—plus one point for each trick won. For example, if a player bets that they will win one trick—and wins one trick—they get one point for winning the trick, but 10 points for betting correctly, for a total of 11 points. Once all tricks are completed, the player with the most points wins.

This game heavily leverages players’ IC and CF; they must be strategic yet calculated in their betting, but they must also be flexible when the tricks within the rounds of the game do not unfold as planned (Miyake et al., 2000). WM is tapped as players’ hold in mind the rules, their bets, as well as the cards that have been played, combining and applying this information fluidly (Diamond, 2013). Finally, as discussed above, players must use ER to keep frustration in check.

***Association***

Association is a rule-based game where players must come up with words that match a rule determined by the leader of the game. Players must use their WM to keep in mind the words that have already been offered, as well as CF, IC and ER. Rules can be category based (ie. types of chocolate bars), rhyme based (ie. words that rhyme with “rock”) or more complex rules, such as words that contain a certain letter. This game can be played with a whole classroom; however, it may be best to split into smaller groups once the game is taught to keep the game faster paced. BBT offers multiple versions of this game to provide a scaffolded challenge for players as well. Game play starts with the announcement of the rule and then proceeds around the circle; each player offers a word they think fits the rule. In the event the word doesn’t fit, is a repeated word, or the player can't think of a word within about five seconds, they lose a life. The game continues until only one player remains.

As mentioned, players must actively use their WM to remember what's been said to avoid repeating words (Diamond, 2013), but they also heavily leverage CF, specifically in the more difficult version of the game that requires players to deduce the rule versus being told the rule explicitly. In line with previous games, this game also requires focus and IC as well as ER.

**Rationale: Physical Activity (PA) Games**

PA is important for all ages, however, instilling the importance of PA at a young age can support the value of physical fitness well into the future (de Bruijn et al., 2018; Dishman et al., 2006). Games from the BBT Adolescent curriculum not only include tabletop games but games that require more movement as well. The amount of PA in the games varies, and individual effort would vary even further. However, in a world that is increasingly sedentary, any increase in PA is beneficial. The following games vary from primarily isolating motor coordination skills (ie. Blindfold Lego) up to moderate-to-vigorous PA (ie. Dodgeball), but all continue to exercise EF skills.

***Rock-Paper-Scissors Tag***

Rock-Paper-Scissors Tag is a team-based variation of the traditional school-ground tag and the game promotes the use of WM, CF, IC and ER. The two teams secretly decide amongst themselves if they will choose rock, paper or scissors (or any other variation that is preferred) and then line up one step away from the middle of the assigned space. The leader counts down, saying "3, 2, 1…SHOW," and each team turns into the middle and displays their choice. The team with the "winning" hand (ie. rock beats scissors) then chases the other team back to their base. Anyone tagged before reaching the base then joins the "winning" team.

This game employs CF and IC as players literally have to think on their feet. They need to make plans and judgements, and react on a moments' notice (Miyake et al., 2000). They also have to use WM to keep in mind the last few choices of two teams to help inform their future plans (Diamond, 2013). Finally, ER allows players to choose their response in challenging situations (Poon, 2018; Zelazo & Carlson, 2012).

***One Word Story: With a Ball***

One Word Story: With a Ball is an adaptation of the original "One Word Story" game, where players take their turn by going around in a circle. In this version, players pass their turn by passing a ball. This game heavily uses WM and CF but also employs IC. This game can be played with classroom-sized groups but is faster-paced with smaller groups. The goal of the game is to complete a full story with each player only adding one word at a time. The first player starts the story with the second player only adding one word (ie. "Once"..."Once upon…"). Play continues until a player makes a mistake (ie. grammar or syntax, such as: "Once upon frog"), can't think of a word, or drops the ball. Leaders can vary the difficulty of the game by either having players only add one word, or by having players repeat the whole story, then having them add their word. The latter requires a great deal of focus and WM as they remember and then manipulate the information (Diamond, 2013). Players also need to use CF; by using the ball to pass the turn, turns aren't automatic, and they must attend to the details and be prepared to add a word to the story at any moment. This also requires IC as they can't speak unless it's their turn.

***Blindfold Lego***Ⓡ

Blindfold Lego is a game that builds on a game from the BBT's original preschool curriculum (Coelho et al., 2020) and leverages the connections between MC, EF and communication. Two players are paired up, one is the blindfolded builder, and the other player is the eyes. The player that's not blindfolded (The Eyes) can see the model as well as the Lego in front of the builder; the builder must build a replica of the model as described by The Eyes. As The Eyes can see the bricks and the model, they must describe to the builder what they are building, how to build it, and where to find the pieces. While this game relies heavily on communication skills, it also requires both parties to plan, make reasoned judgements and to inhibit distractions or the impulse to simply use any piece (sometimes there are multiple 2x2 pieces, and the builder needs the eyes to tell them which one to use to replicate the model exactly). In addition to this, the removal of a key sense (vision) is frustrating, and often players need to use ER to get through the process of assembling more challenging models (Poon, 2018; Zelazo & Carlson, 2012).

***1 Tap, 2 Tap***

Similar to Blindfold Lego, 1 Tap, 2 Tap involves more controlled and precise motor coordination. It also leverages WM, IC, and ER. To play 1 Tap, 2 Tap, players lay down on their stomachs in a circle, with their heads and hands towards the middle. Each player criss-crosses their arms with other players' arms so that each player has one arm over, and one arm under the players to their left and right. The first player indicates the direction of travel and taps their hand on the ground; play follows in the order of hands, this means that player one might start the round, but player three taps second because their hand is next in the circle of hands. Play continues, but the direction of travel can be reversed by a player tapping twice, as opposed to once. Play can be quite strategic as players work to "trap" others' hands. When a player makes a mistake, they remove that hand from the circle; they are eliminated once they lose both hands. The last player "standing" wins.

While it might sound simple to keep the game rules in mind, once players are using both cognitive and physical resources it can become quite challenging, so WM is crucial (Diamond, 2013). IC is also important in this game as players need to focus, select the correct hand and inhibit the incorrect motion (Miyake et al., 2000). Finally, as in most games, ER is important so that the youth can continue to engage appropriately with their peers despite any setbacks.

### *Dodgeball*

Dodgeball is a game that requires moderate-to-vigorous activity regardless of the variation being played and it generally has a team component as well. While there are many variations of the game, the general rules include boundaries, what happens when a person is hit or catches a ball and how players can re-enter the game. Dodgeball largely relies on CF and ER. Players need to be fluid in a very dynamic environment and need to shift their attention between multiple stimuli. Dodgeball can also produce emotional dysregulation, so learning to build and control ER is imperative to be able to continue play successfully (Poon, 2018; Zelazo & Carlson, 2012). Finally, Dodgeball requires WM, especially versions that require players to attend to, and remember, which player tagged them so they can reenter the game once this player is tagged themselves (see "crackabout" from Smart Moves, n.d.; Diamond, 2013).

## Conclusion

While there is a variety of literature currently available regarding EF, adolescent EF is understudied in comparison. In order to fill this gap, the BBT adolescent curriculum was developed (Fleischman et al., 2022) and studied to determine if improvements similar to the 2021 study conducted by Gibb et al. can be seen in a teenage population. Chapters 3 through 6 discuss the procedure and findings of this study and provide support for a play-based intervention designed to support adolescent EF.

# CHAPTER 3: METHODS

## Research Design

This study in question was structured as a quantitative quasi-experimental design, with the intervention, and participants' individual dosage, being the independent variable, and the battery of pre- and post-tests being the dependent variable. In addition to these variables, a number of control and moderating variables were collected on the initial demographic sheet, including age, gender, and handedness. This study does not use randomly assigned control and experimental groups as the research took place in a school where the classes were already established. Exclusionary criteria for participation in the study included a lack of availability for testing (ex. student wasn't present on any of the testing days), significant cognitive impairments (ex. Autism Spectrum Disorder, severe cognitive delays, etc.) and refusal of assent (i.e. the student declined to participate).

### *Participants*

Participants included two classrooms of Grade 7 adolescents from a participating middle school in Lethbridge, Alberta. While the groups themselves were not random, one classroom was arbitrarily assigned to be the experimental group *(n* = 31), and the other to the control group (*n* = 30); the intent was to have the control group participate in the testing after the first experimental condition was completed. Both classes were from the same middle school, and both groups share the teachers involved in the experiment (i.e. the two classrooms were dyads, with the classes switching between the two teachers for humanities and sciences). Total participants analyzed in the experimental group was *n* = 27; four participants were excluded from data collection due to cognitive impairments or lack of availability for testing. Total participants analyzed in the control group was *n* = 29. This brings the total to *n* = 56 (22 female; see Figure 1 for demographics).

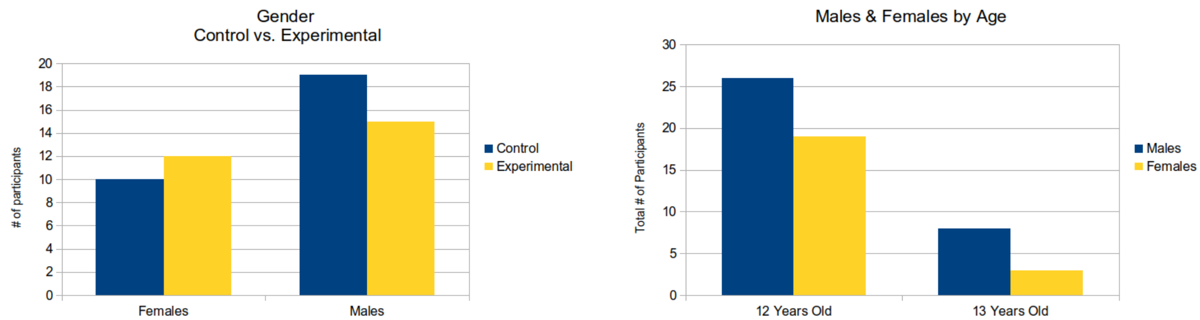


Figure 1: Gender by Treatment (left) and Participant Age (right)

The original experimental design included having the control subjects participate in the intervention after post-testing had concluded to allow for both between-subject experimentation and within-subject experimentation. However, due to the realities of action research in schools, the timeline did not permit us to conduct another round of experimental testing in the school year.

***Materials***

Participants engaged in 6 weeks of the BBT Adolescent Curriculum (Table 1) at a dosage of three days per week, for approximately 15 minutes per day. The games were delivered during the school day by the classroom teacher.

**Table 1**

*Building Brains Together Adolescent Curriculum*

| Game | Description | Executive Functions Targeted |
|---|---|---|
| 1. Spoons | Card game, quick reaction elimination-based game | Working memory, inhibitory control, emotional regulation |
| 2. Slap | Card game, quick reaction elimination-based game | Working memory, cognitive flexibility, inhibitory control, emotional regulation |
| 3. Rock, Paper, Scissors Tag | Physical movement game, team versus team rock, paper, scissors | Working memory, cognitive flexibility, inhibitory control, emotional regulation |

| | | |
|---|---|---|
| 4. One Word Story | Physical movement game, create and repeat a story by adding one word at a time | Working memory, cognitive flexibility, inhibitory control |
| 5. Concentration | Verbal game, silently pass the turn through the group by using either numbers or symbols. | Working memory, cognitive flexibility, inhibitory control, emotional regulation |
| 6. 1 Tap, 2 Tap | Physical movement game, using taps, change the direction of the flow of the game to eliminate others | Working memory, inhibitory control, emotional regulation |
| 7. Oh Heck | Card game, players bet on the round and each gets points based on how close to correct they were | Working memory, cognitive flexibility, inhibitory control, emotional regulation |
| 8. Blind Fold Lego | Tabletop game, players are blindfolded and guided through building a Lego figure by a partner | Inhibitory control, emotional regulation |
| 9. Association | Verbal game, a leader devises a "rule" and players must guess the rule before their lives run out | Working memory, cognitive flexibility, inhibitory control, emotional regulation |
| 10. Dodgeball / Gaga Ball | Physical movement game, players attempt to get each other "out" by tagging each other with a ball | Working memory, cognitive flexibility, emotional regulation |

***Procedures***

Participants underwent two rounds of testing, pre- and post-intervention. Pre-testing (T1), was the initial testing of participants prior to intervention. Post-testing for the experimental group (T2) occurred after 6 weeks of the delivery of the curriculum of games. The time period between T1 and T2 was 12.5 weeks on average for the experimental group and 6 weeks on average for the control group.

The battery of tests included the BRIEF-SR survey, five psychometric tests and two Lego brick building tasks (LBBTs), all of which are outlined below; each test assessed one or more EFs or related measurements, as explained in Table 2. Four of the psychometric tests (List Sorting Working Memory Test, Dimensional Change Card Sort Test, Flanker Inhibitory Control and Attention Test, and Picture Vocabulary Test) were completed using the National Institutes of Health (NIH) Toolbox and the fifth psychometric test (Stroop) was completed using "The Experiment Factory" (Sochat, 2018) accessed via a browser.

**Table 2**

*EFs Being Assessed*

| Test | Assessed via | Measurements Collected | EFs Assessed |
|---|---|---|---|
| BRIEF-SR | Paper survey | Behaviour Regulation Index (BRI), Metacognition Index (MI) and General Executive Composite (GEC; BRI + MI)[1] | Emotional regulation[1]<br>Inhibitory control[1]<br>Working memory[1]<br>Cognitive flexibility[1] |
| Stroop Test | The Experiment Factory | Congruent response time (RT)<br>Incongruent RT<br>Average RT<br>Congruent score<br>Incongruent score<br>Average score | Selective attention[2]<br>Inhibitory control[3]<br>Cognitive flexibility[3] |
| List Sorting Working Memory Test | NIH Toolbox | Age-corrected standard score | Working memory[4] |
| Dimensional Change Card Sort Test | NIH Toolbox | Age-corrected standard score | Cognitive flexibility[5] |
| Flanker Inhibitory Control and | NIH Toolbox | Age-corrected standard score | Inhibitory control[5] |

| | | | |
|---|---|---|---|
| Attention Test | | | |
| Picture Vocabulary Test | NIH Toolbox | Age-corrected standard score | Receptive language[6]<br>Vocabulary[6] |
| Lego brick building tasks (LBBTs) | In-person motor task | Build time<br>Space use<br>Handedness<br>Build errors | Spatial perception[7]<br>Spatial visualization[7]<br>Mental rotation[7] |

*Note.* [1]Walker & D'Amato, 2006
[2]Alvarez & Emory, 2006
[3]Stuss et al., 2001
[4]Tulsky et al., 2014
[5]Zelazo et al., 2013
[6]Gershon et al., 2013
[7]Aguilar et al., 2021

The NIH tests were counterbalanced by creating four separate test batteries each with a different order of tests (see Table 3). Administration of the Stroop test was counterbalanced by alternating whether participants completed it before, or after, the NIH tests. The LBBTs (2D and 3D models) were counterbalanced against the above psychometric tests by alternating whether participants completed the building task first—before the psychometric tests—or after them. They were also counterbalanced within the task by alternating whether participants completed the 2D models or 3D models first.

**Table 3**

*Counterbalancing of NIH Toolbox*

| BBT/C4L-1 | BBT/C4L-2 | BBT/C4L-3 | BBT/C4L-4 |
|---|---|---|---|
| 1. PVT<br>2. Word list WM<br>3. Card sort<br>4. Flanker | 1. Card sort<br>2. PVT<br>3. Flanker<br>4. Word list WM | 1. Word list WM<br>2. Flanker<br>3. PVT<br>4. Card sort | 1. Flanker<br>2. Card sort<br>3. Word list WM<br>4. PVT |

***BRIEF-SR.*** Participants were provided the BRIEF-SR by either the researcher or the classroom teacher and given the instructions to consider their behaviour over the previous 6 months. Students were given the option to read the survey questions themselves, or to have the researcher or teacher read the questions to them. During pre-testing, when the BRIEF-SR was conducted by the researcher, only one participant requested that the questions be read aloud to them.

***Stroop***. The Stroop color-word test was conducted using the open-source software available on The Experiment Factory website (Sochat, 2018). The test was run using the Safari or Chrome browser, and the results were exported and saved locally. The Stroop test consisted of 18 practice items followed by 72 trial items. The test trials were randomly generated and included 48 congruent items (word = BLUE and the word font color is BLUE), and 24 incongruent items (word = BLUE, but the word font color is either yellow, or red). See Figures 2 and 3 for examples.

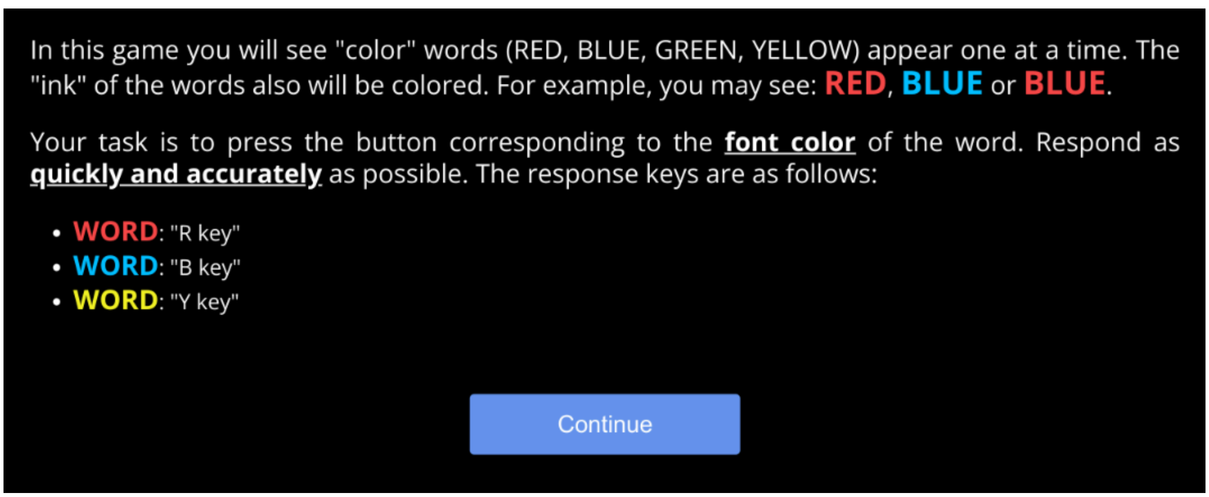


Figure 2: Stroop Test Instructions (Sochat, n.d.)

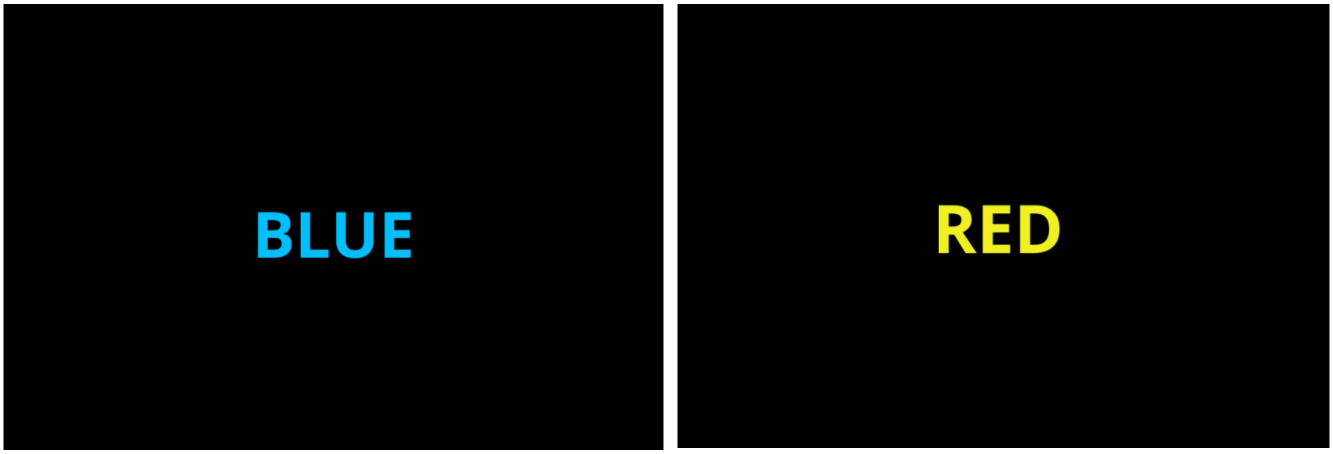


Figure 3: Examples of Stroop Test Items

*Note.* Congruent test item (ink and word are the same colour) is on the left, and an incongruent test item (ink and word are different colours) on the right (Sochat, n.d.).

***List Sorting Working Memory Test (LSWM).*** As per the NIH Toolbox App Administrator's Manual v1.23 (National Institutes of Health, 2020), the LSMW test was overseen by a researcher who inputted the participant's responses. In this test, there were two separate but related trials. For the first trial, participants were given a list of words (either animals, or food), while simultaneously seeing the image that corresponds with the word presented on the tablet, and were instructed to repeat the items back to the researcher in size order, from smallest to largest (see Figure 4 for an example image). For the second trial, participants are given a list that includes both food, and animals, and were again asked to repeat the items back to the researcher in size order, from smallest to largest, but with the additional instruction to sort the food items first, and the animals second. In both trials, the number of items presented gradually increased from the initial two-items to a maximum of seven-items. Participants were given two practice items prior to the initiation of each trial.

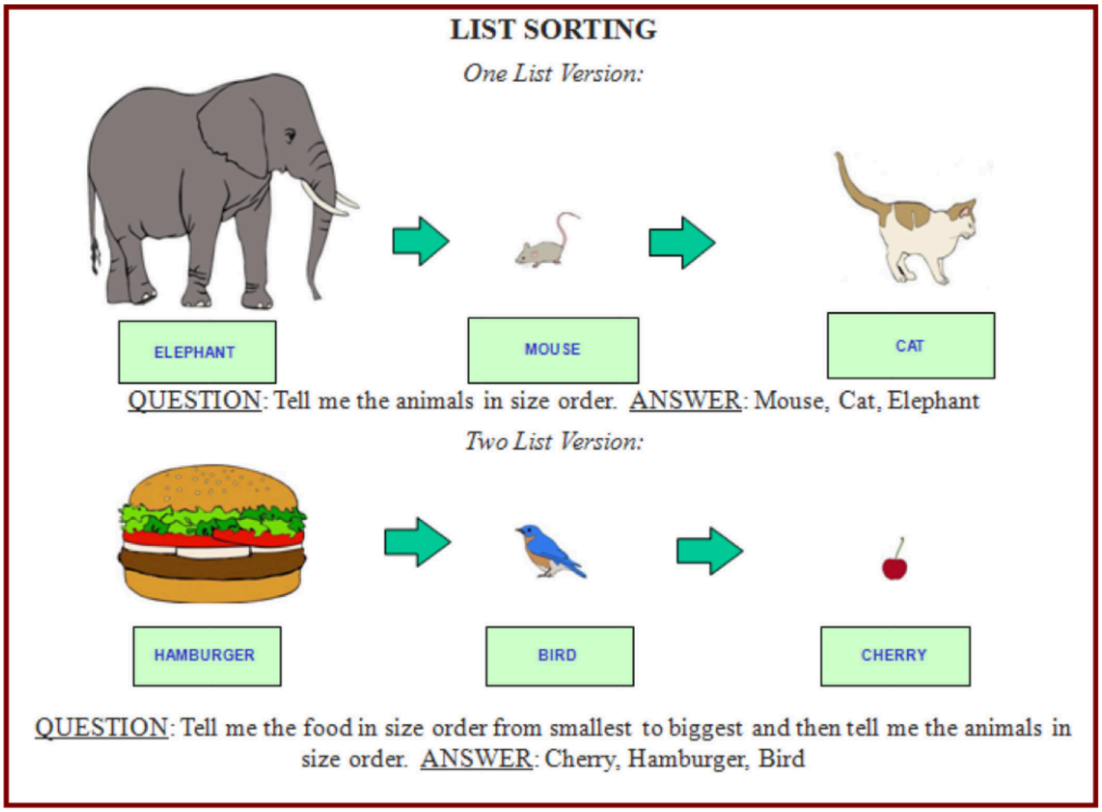


Figure 4: Image explaining the LSMW Task in the NIH Toolbox (Tulsky et al., 2013)

***Dimensional Change Card Sort Test (DCCS)***. The Dimensional Change Card Sort Test (DCCS) was supervised by a researcher, however, it was presented and run by the NIH Toolbox application. The DCCS presented the participant with two images (ex. a blue truck and a yellow ball) and an initial sorting rule (ex. Sort by SHAPE, or sort by COLOUR; see Figure 5 for an example). The scored task was preceded by two practice items. After the participant finished sorting by shape, the test changed the rule and instructed the participant to sort by colour; again, preceded by two practice items. Finally, the test presented a combined trial, where participants needed to employ inhibitory control and flexible thinking to respond to the trial's randomized prompt to sort by either COLOUR or SHAPE.

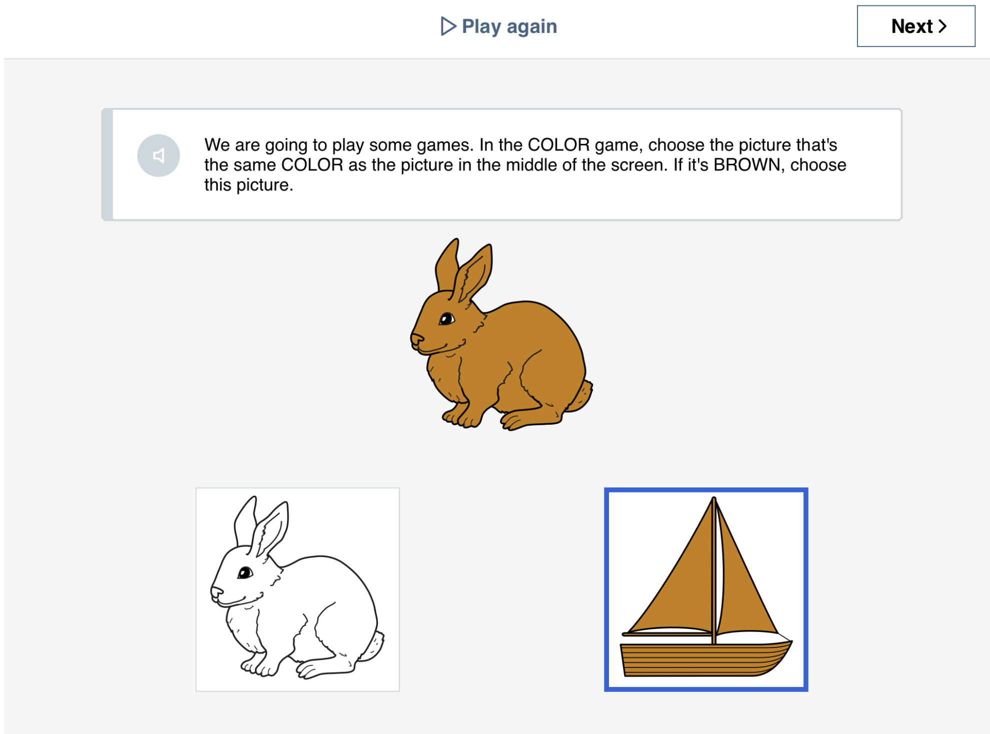


Figure 5: Example of the DCCS task in the NIH Toolbox (NIH Toolbox, n.d.-a)

***Flanker Inhibitory Control and Attention Test (Flanker).*** The Flanker Inhibitory Control and Attention Test (Flanker) required the participant to both focus on a stimulus (the arrow in the middle of a series of arrows; pointing either left or right) and inhibit attention on the flanking arrows (see Figure 6 for an example). Based on the direction of the stimulus, the participant then had to select whether the stimulus arrow is pointing left or right in a rapid-answer format.

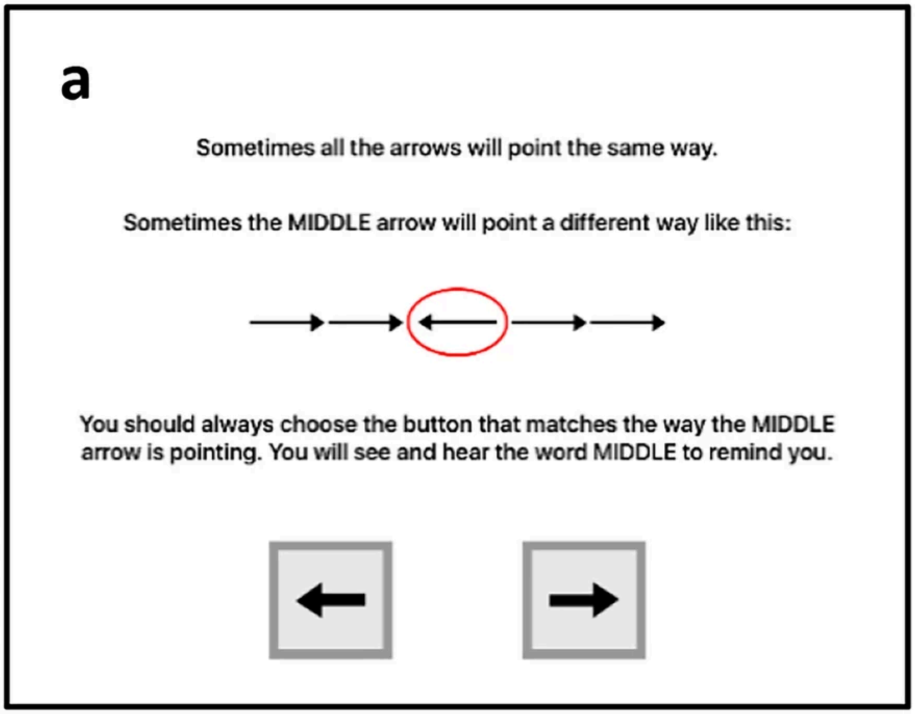


Figure 6: Example of the Flanker Task on the NIH Toolbox

*Note.* Image adapted from Jones et al. (2022).

***Picture Vocabulary Test (PVT).*** The Picture Vocabulary Test (PVT) is the NIH Toolbox's variation on the Peabody picture vocabulary test. It presented the participant with a word, narrated orally through the application, while it concurrently presented the participant with four images (see Figure 7). The participant had to select the image that correctly demonstrated the provided oral word, in the case of Figure 7 "spoon." The application adapted the difficulty level as the participant answered correctly or incorrectly.

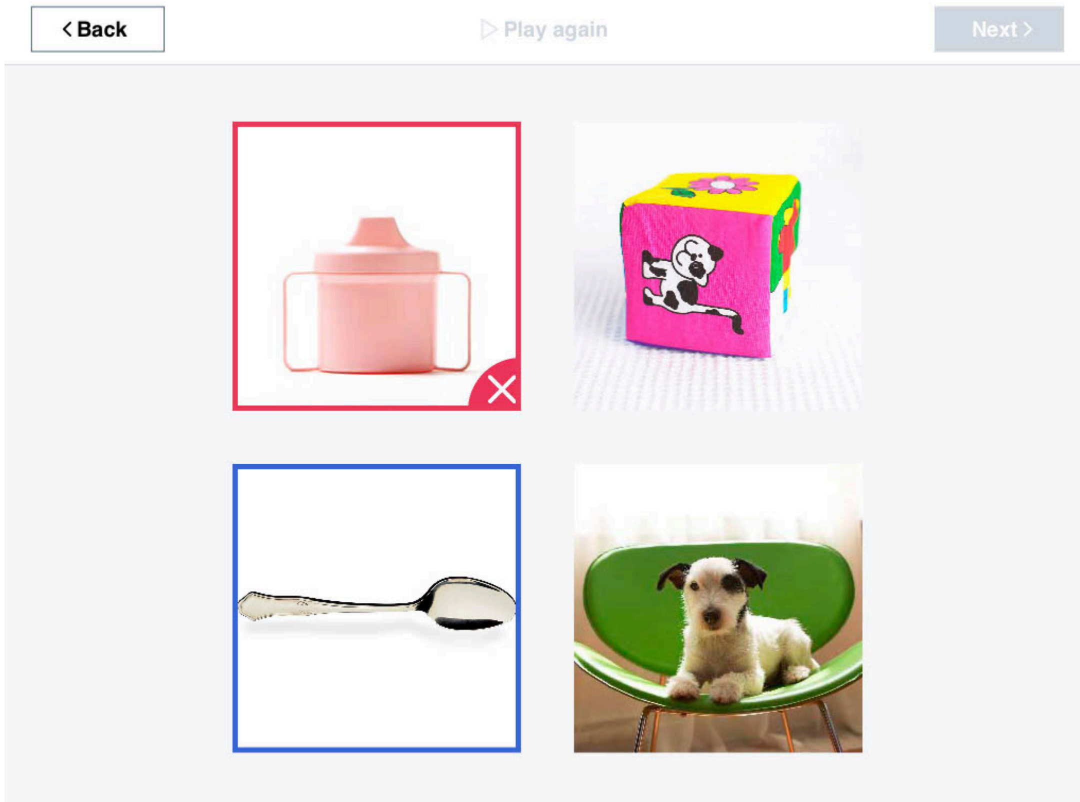


Figure 7: Sample of the PVT on the NIH Toolbox (NIH Toolbox, n.d.-c)

***Lego Brick Building Task (LBBT).*** The LBBT, devised by the lab of Dr. Claudia Gonzalez, was designed to assess the handedness of participants, fine motor skill, executive function and spatial ability. There were two rounds of testing to complete this task: one with "2D" models, and the other with "3D" models (see Figure 8 for visual examples). Each round had three models for a total of six models (three 2D and three 3D). For each round, 36 bricks were laid out on a surface in such a way that the pieces were evenly distributed between the left and right sides. Each model was made of 12 pieces. Therefore, all the bricks were used if participants built the three models correctly. Participants were given a visual model, and instructed to recreate it exactly, and as quickly as they could. The researcher timed their attempt and made note of any errors. The trials were video recorded to be scored after the fact. Once the participant finished the first round, they were asked to leave the room, or turn around, while the researcher set up the next brick task. The participant then completed the second round, which was determined in advance by counterbalancing.

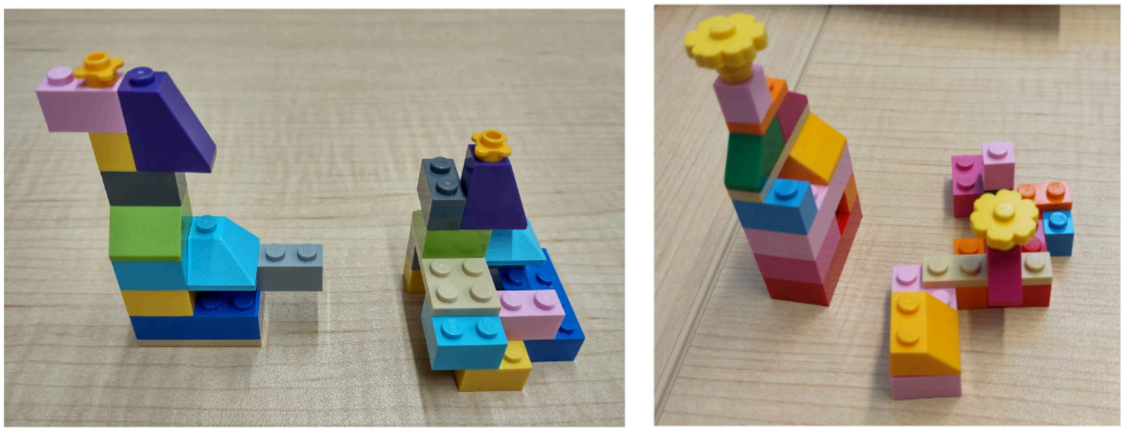

Figure 8: 2D and 3D Brick Models for the LBBT

*Note.* The above are two of the three sets of models presented to participants. In both images the "2D" model is pictured on the left and the "3D" model is on the right.

**Data Collection**

Data was collected at two time-points, with pre-testing done between January 30, 2023 and March 15, 2023, and post-testing completed between April 18, 2023 and May 8, 2023. When at all possible, participants were tested in the same order during both time-points. During both pre- and post-testing, data was collected via the BRIEF-SR survey, the NIH Toolbox (delivered via an iPad), and through a video-recorded brick-building motor-task.

***BRIEF-SR***

The Behavioural Rating Inventory of Executive Functions, self-report version (BRIEF-SR) is an 80-item survey that seeks to capture a youth's perspective on their own EF abilities. The series of BRIEF surveys were developed by Dr. Steven Guy, Dr. Peter Isquith, Dr. Gerard Gioia and Dr. Kenworthy (Isquith et al., 2013) with the purpose of closing a gap in neuropsychological scales by providing a much needed measurement tool for executive function.

The survey takes participants 10-15 minutes to complete, and requires them to answer “never,” “sometimes” or “often” to the 80 items. To support the validity of the instrument, the survey was written in such a way that allows the researcher to track both “negativity” and “inconsistency.” Reviews of the instrument, and the professional manual itself, detail the thorough norming of the instrument along with its reliability and validity. When the survey is scored, the scores are grouped into eight categories: inhibit, shift, emotional control, monitor, working memory, plan/organize, organization of materials and task completion. After calculating these raw scores, the first four categories (inhibit, shift, emotional control, and monitor) were added together to obtain the Behavioural Regulation Index (BRI) and the latter four scores (working memory, plan/organize, organization of materials and task completion) were added together to obtain the Metacognition Index (MI). Together, the BRI and the MI create the Global Executive Composite (GEC). Once these raw scores were computed, they were converted to T-Scores using the appendix in the manual. The GEC (T-Score) was the score used in this study as it provides the full picture of a participant’s EF skills and is the most consistent score on the BRIEF, as noted below.

The BRIEF-SR is commonly used in conjunction with the BRIEF parent or teacher form in order to provide both an internal and external measure of a participant’s EF. However, due to the complicated nature of working in a school, this study was only able to use the BRIEF-SR. The BRIEF-SR is reported to correlate well with the adult survey (Egan et al., 2019; Walker & D’Amato, 2006). According to Walker and D’Amato (2006), “Internal consistency using Cronbach’s alpha was reported for the eight clinical scales, the two broad indexes, and the overall GEC. The internal consistency was high for the normative sample for the 80-item GEC

(.96) and moderate for scales with fewer items (.72)" (p. 397). Furthermore, test-retest, and interrater reliability were both moderate-to-high.

### *NIH Toolbox*

This study used versions 1.23-1.25 of the NIH software, and version two (V2) of the app, throughout the duration of the testing. It is worth noting that in early 2023, NIH released V3 of the app, however, due to a lack of compatibility between V2 and V3, the study continued to only use V2.

The NIH toolbox is a robust measurement tool that provides normed tests in the domains of cognition, motor, sensation and emotion. In order to use the tests in the cognition section, additional approval is required as per the APA's Standards for Educational and Psychological Testing, and after applying for approval, permission was granted by the NIH Toolbox.

Four tasks were used within the NIH Toolbox, these included: the NIH Toolbox Dimensional Change Card Sort (DCCS), the NIH Toolbox Flanker Inhibitory Control and Attention Test (Flanker), the NIH Toolbox List Sorting Working Memory Test (LSWM) and the NIH Toolbox Picture Vocabulary Test (PVT). Norming of the instruments within the toolbox included a "sample of 4,859 participants, ages 3-85, representative of the U.S. population based on gender, race/ethnicity, and socioeconomic status" (National Institutes of Health, 2020).

### *Stroop Test*

The Stroop test is a well-known and validated psychometric test that uses the go/no-go framework to ascertain the participant's ability to inhibit their prepotent, or automatic, response, in favour of the correct response. This is done through the paradigm of colours and colour-word combinations where participants must identify the colour of the ink a word is written in; the difficulty of this task stems from the alternation between congruent (ex. The word "BLUE" written in *blue* ink) and incongruent (ex. The word “BLUE” written in *red* ink) trials. The Stroop test measures response times for congruent and incongruent trials, as well as the participants’ correctness within both of those trials.

For this study, the Stroop test was administered using an online, open-source program provided by the Poldrack Lab at Stanford University (Experiment Factory; Sochat et al., 2016). The test in question was facilitated through a browser, with results being immediately downloaded and saved to the iPad dedicated to the study. The data provided by the program included practice and test items, as well as stimulus and fixation items; therefore, the data needed to be thoroughly cleaned before it was analyzed.

The data was manually processed by the researcher and one assistant through spreadsheet manipulation. By selecting for certain keywords (fixation, practice) and durations (500ms for fixations), the researcher was able to eliminate all extraneous data lines. Once these data were removed, simple addition and average formulas were used to calculate congruent and incongruent response times, as well as correct and incorrect trials of each.

***Lego® Brick Building Task (LBBT)***

The LBBT was designed to test a variety of factors in children and youth including spatial perception, spatial visualization and mental rotation (Aguilar et al., 2021). Furthermore, this task was created to provide an age-appropriate alternative to the mental rotation test. Tested with children ages 5-17, for this study, the task was chosen to provide additional information regarding participants’ handedness, spatial abilities and ultimately their EF skills.

There are two trials within the LBBT, the 2D trial and the 3D trial. The order of these trials was alternated for pre- and post-testing to counterbalance the task. These tasks were video recorded in order to be scored later, but were angled in such a way that only the blocks and the participants' hands were in view to maintain anonymity of the participant. Regardless of which trial the participant started with, they began with the same 36 bricks set out equally, but randomized, in four quadrants, for a total of nine bricks per quadrant. These 36 bricks would be fully utilized to build the three models in the trial. Due to the nature of conducting testing in a school environment, the space utilized for the testing varied slightly, as did the surface used for the LBBT. The LBBT was set up in a rectangular fashion, ranging from an area of 30-by-24 inches (i.e. a standard pupil desk), to an area of no more than 76-by-60cm (36-by-24 inches). When the bricks were distributed between the four quadrants, the researcher left a vague midline that was only identifiable by the research team to help with post-hoc scoring.

The LBBT was either set up before the participant entered the room, or the participant was asked to turn around while the researcher set up the task to prevent any preemptive identification of pieces. The models were also hidden from view until immediately before the participant was to start the trial. Participants were given the instructions to replicate the researcher's model as quickly, but as accurately, as possible–the same shape, colour and size. They were also instructed that they could pick up the model and review it at any time, they simply could not deconstruct it. Participants were asked to start when the researcher said "go," and to verbally indicate when they felt they were done building the model; they were timed between these points. Researchers also noted any errors participants made in either: selecting the wrong block, or wrong block colour, and incorrect block placements. These errors were noted both during the trial, as well as retrospectively when scoring the videos offline.

**Data Analyses**

Both the control group and experimental group were tested both pre- (T1) and post-curriculum (T2) delivery using the battery of tests listed in procedures. Analysis included reviewing the relationship between T1 & T2 in the individual subjects as well as comparing the experimental to the control group.

All statistical analyses were conducted using IBM SPSS Statistics, version 29.0.2.0 (20), licensed through the University of Lethbridge. Prior to analysis, data was cleaned and reviewed for accuracy. The BRIEF-SR, Stroop test and the LBBT were cleaned or scored by the researcher, however, the NIH toolbox automatically provides the age-corrected standard score (among others) for the researcher's use as pictured in the output below (Figure 9).

| PIN | Assessment Name | Inst | DateFinished | Uncorrected Standard Score | Age-Corrected Standard Score | National Percentile (age adjusted) | Fully-Corrected T-score |
|---|---|---|---|---|---|---|---|
| ABF13 | Assessment 2 | NIH Toolbox List Sorting Working Memory Test Age 7+ v2.1 | 2023-04-28 9:29:28 | 128 | 139 | 99 | 72 |
| ABF13 | Assessment 2 | NIH Toolbox Flanker Inhibitory Control and Attention Test Age 12+ v2.1 | 2023-04-28 9:32:45 | 111 | 111 | 77 | 59 |
| ABF13 | Assessment 2 | NIH Toolbox Picture Vocabulary Test Age 3+ v2.1 | 2023-04-28 9:34:26 | 98 | 112 | 79 | 63 |
| ABF13 | Assessment 2 | NIH Toolbox Dimensional Change Card Sort Test Age 12+ v2.1 | 2023-04-28 9:39:05 | 118 | 142 | 99 | 75 |

Figure 9: Example of NIH Toolbox Output

*Note.* Sample from ABF13 post-test. More information is provided in these outputs (raw score, app version etc.), however this CSV file has been cleaned for illustrative purposes.

Following this collection, the data was collated into a master spreadsheet that included both the control and experimental groups. This data was transferred into SPSS's data view for analysis with a total of 11 variables for each pre- and post-condition. As previously mentioned, the NIH Toolbox tasks were outputted as a single standard score, and once calculated the BRIEF-SR provided a standardized T-Score called the GEC. The Stroop produced six different variables on one of two dimensions: response time (RT) or score. Furthermore, the Stroop test provides data on the participant's results on both congruent (word reads "blue," and font is blue) and incongruent (word reads "blue," but font is red) trials. Accordingly, for the score dimension, variables include a total score, a score for incongruent trials, and a score for congruent trials, all entered as percentage answered correctly. For the response time dimension, there is an overall average RT, an incongruent trial RT and a congruent RT, all entered in seconds.

The descriptive statistics were reviewed in SPSS for skewness, kurtosis, tests of normality, and the test of homogeneity of variance. This also included reviewing the dependent variables for outliers using both the SPSS “explore” function, as well as manually reviewing z-scores for all the variables. Building on this, the descriptive statistics were bootstrapped ($n$ = 1000) to check for any changes in normality as the data set is limited with only $n$ = 56. Following the results of the tests of normality, the analyses branched off into both parametric and non-parametric analyses. For all analyses, when appropriate, cases were excluded pairwise to allow for the inclusion of the most data possible, given the smaller data set.

Non-parametric tests that were included in the analyses were the Wilcoxon signed-rank test, the Mann-Whitney U test, and the Friedman test. While these will be further discussed in the results section, these tests were used to get a full picture of both the within-subjects’ effects, as well as the between-subjects effects. These tests were only used with the dependent variable that were found to violate parametric assumptions.

Parametric tests conducted included paired-samples t-tests, and both multivariate and repeated measures (RM) general linear models (GLM). Relationships explored in the GLMs primarily included the interactions between time, gender, treatment and dosage. Building on the GLMs, a basic general linear regression was employed to look at the impact of dosage on the changes in participants' scores.

## Methodological Limitations

### *Quasi-Experimental Design*

While a full randomized control trial is the ideal model, given the nature of the population and location that we sought to work with, in order to test an intervention that was designed to be functionally implemented in a classroom, a quasi-experimental design was the best choice for this study. While the classrooms were randomized at the beginning of the school year, there was a chance that there may be some unanticipated variation between the groups. However, based on descriptive statistics and demographic information, there didn't appear to be significant differences. In fact, in reviewing mean pre-scores, the greatest difference between the two treatments was a variation of 4.33% on the Stroop Incongruent pre-score with the control group averaging 77.59% correct and the experimental averaging 74.23%. The DCCS and PVT had similar variation with 3.35% and 3.01% respectively (see Figures 10 and 11 for a pre-score comparison by treatment and Table 6 for detailed figures).

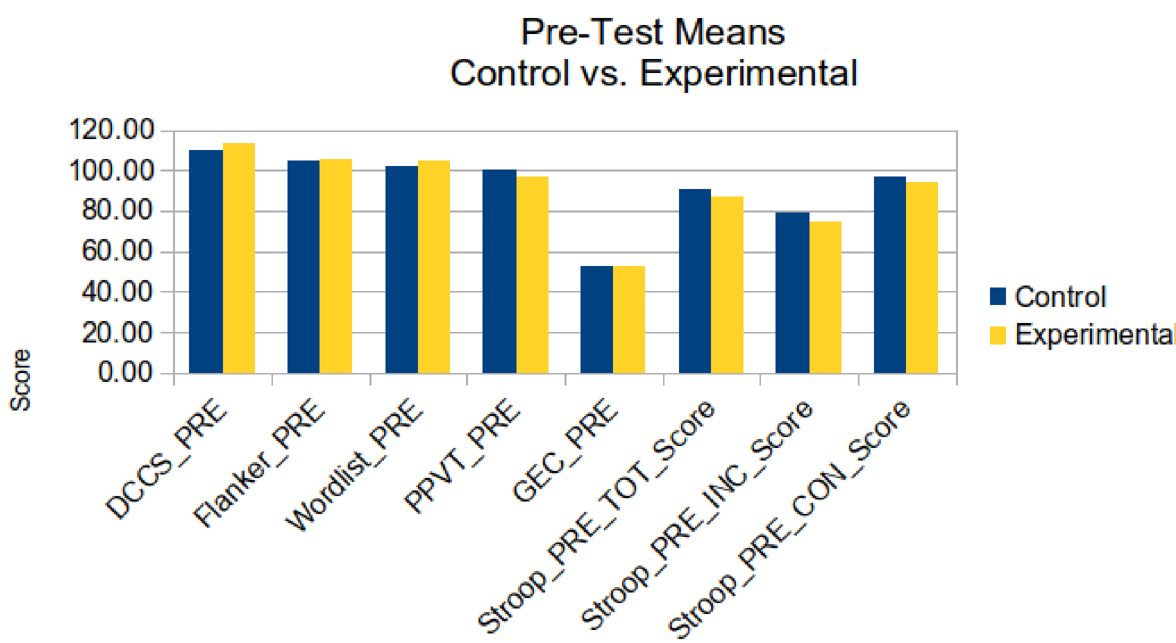


Figure 10: Pre-Test Scores by Treatment

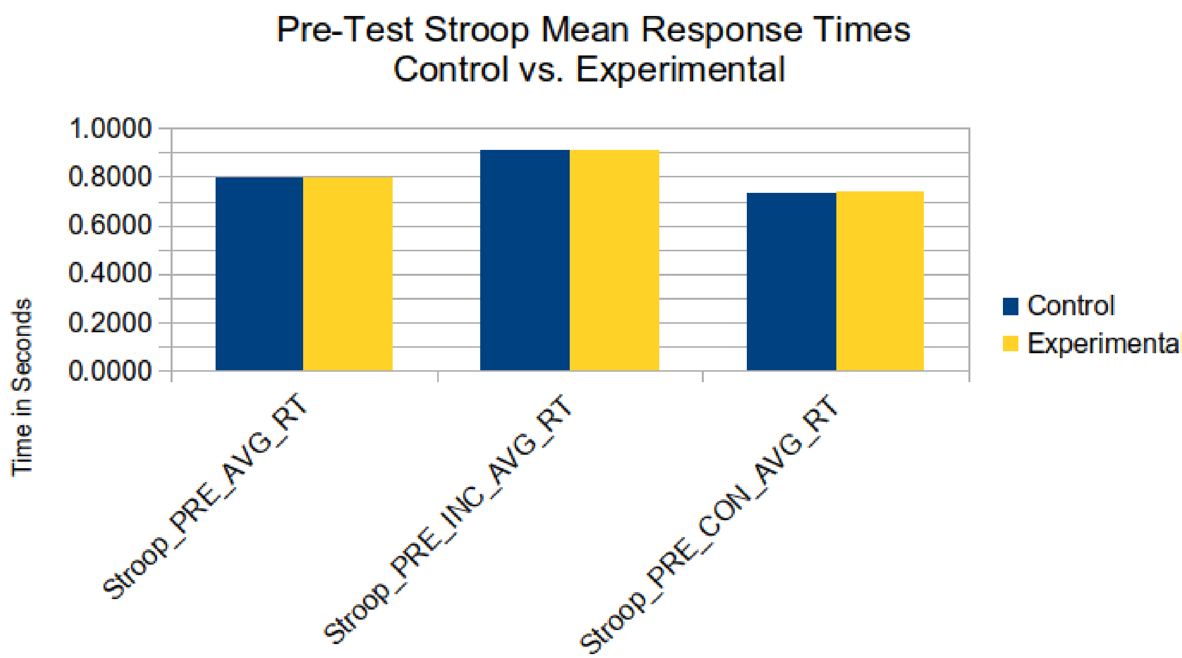


Figure 11: Pre-Test Stroop RT by Treatment

It was decided at random which classroom would begin with the intervention, and the original plan was for the control group to also undergo the intervention after T2. However, due to timing and school variables, this was not possible.

### *Self-Selection Bias*

Another potential limitation was the potential for participants to self-select themselves out of the testing. However, none of the participants in either the control or experimental group opted-out. The only participants that were not included in the analyses were those that were behaviourally or cognitively unable to complete the testing protocol, or those whose were absent for the duration of the testing.

### *School-Based Experimental Testing*

One of the challenges of school-based research is the dynamic and fluid environment. Unlike a designated research lab where materials can be organized, locked-up and controlled, conducting testing in a school requires flexibility, mobility and a high-degree of collaboration. However, also unlike a research lab, students are more likely to be at ease in a familiar environment, so while there were drawbacks to testing within the school, there were considerable advantages as well.

Testing during this study was conducted in one of four locations, two of which were the primary locations, and were used the most often. The first space was a small conference room that was used at the onset of the testing while the administration found a more suitable and sustainable location. The second and third locations were spare offices that could be booked by contract staff, such as counselors; these rooms were used most often and worked well when two researchers were present as the rooms were side-by-side. The fourth location was a small room on the second floor of the school; this room was a somewhat awkward size and layout, but was only used during some post-tests when all three of the other locations were already being used.

All locations allowed for the researcher and the participant to complete all previously mentioned testing components with room to move around as needed. The biggest challenge that was present with all locations was the intermittent background noise. Generally the noise was minimal and non-distracting, but there were occasions where, in location two and three, there was a minor disturbance out in the hallway by the rooms and that distracted participants. While events to this extent were rare, there were often noises associated with classes moving around the school and class changes. Researchers tried to avoid running the testing near class change times, but this was not always avoidable. Therefore, this did add a confounding variable that was not accounted for in the data.

Another aspect of testing in the school that presented some challenges and limitations to the study was the need to be considerate of the class periods throughout the day. It was largely pre-arranged with the cooperating teachers and administrators which classes students could, and could not, miss while participating in testing. However, there were occasions where participants expressed the desire to return to a particular class, even though they may have been excused. Whenever possible we worked with the participants to avoid testing them during their "preferred" classes. The reason for this is that we did not want to add any resentment, or additional distractions, that could potentially impact the testing.

While conducting this study directly in the school had its drawbacks, it also had a number of redeeming features. First, as previously mentioned, the students are in a familiar environment which can help offset the nervousness that can sometimes be present in participants. Second, students and teachers knew the school layout and were able to easily find any of the testing locations. Third, while communication between the researcher and the school team was already positive, working in the school allowed for easy and accessible communication, as well as continued relationship building.

# CHAPTER 4: RESULTS

The goal of this research study was to test the hypothesis that the purposeful inclusion of the BBT adolescent game curriculum would enhance the EF skills of inhibition, working memory and cognitive flexibility in the participants (Sala & Gobet, 2016). While the data provides some support for this hypothesis, further research needs to be conducted to validate the findings. Ultimately, with *n* = 56 participants (22 female), the sample size is limited when attempting to run detailed analyses on multiple variables.

## Descriptive Statistics

The analysis began with a detailed review of the descriptive statistics, with a focus on comparing means, skewness, kurtosis, and the tests of normality—a summary of these results are found in Table 4. For this, and all subsequent tests, the researcher used the pairwise exclusion for any missing data to allow for the inclusion of as much of the data as possible, versus excluding listwise which would have removed entire cases.

**Table 4**

*List of Variables and Related Descriptive Statistics*

| Control | N Statistic | Minimum Statistic | Maximum Statistic | Mean Statistic | Std. Deviation Statistic |
|---|---|---|---|---|---|
| DCCS_PRE | 29 | 68 | 146 | 109.62 | 23.471 |
| DCCS_POST | 29 | 78 | 146 | 116.52 | 21.941 |
| Flanker_PRE | 29 | 76 | 146 | 104.93 | 20.674 |
| Flanker_POST | 29 | 74 | 146 | 109.66 | 21.036 |
| Wordlist_PRE | 29 | 75 | 141 | 101.79 | 18.403 |

| | | | | | |
|---|---|---|---|---|---|
| Wordlist_POST | 29 | 74 | 146 | 104.59 | 17.322 |
| PVT_PRE | 29 | 74 | 135 | 99.93 | 13.630 |
| PVT_POST | 29 | 74 | 126 | 101.66 | 12.619 |
| GEC_PRE | 28 | 32 | 79 | 52.68 | 12.853 |
| GEC_POST | 29 | 31 | 91 | 53.24 | 15.720 |
| Stroop_PRE_AVG_RT (Stroop Pre Average Response Time) | 28 | 0.61 | 1.04 | 0.7933 | 0.11348 |
| Stroop_POST_AVG_RT | 27 | 0.53 | 1.09 | 0.7416 | 0.14486 |
| Stroop_PRE_TOT_Score (Stroop Pre Total Score) | 28 | 76.39 | 100.00 | 90.7250 | 6.13122 |
| Stroop_POST_TOT_Score | 27 | 76.39 | 100.00 | 91.9244 | 5.26767 |
| Stroop_PRE_INC_AVG_RT (Stroop Pre Incongruent Average Response Time) | 28 | 0.65 | 1.24 | 0.9089 | 0.15000 |
| Stroop_POST_INC_AVG_RT | 27 | 0.56 | 1.31 | 0.8268 | 0.17111 |
| Stroop_PRE_INC_Score (Stroop Pre Incongruent Score) | 28 | 41.67 | 100.00 | 79.1671 | 16.27604 |
| Stroop_POST_INC_Score | 27 | 0.75 | 100.00 | 81.3552 | 19.05437 |
| Stroop_PRE_CON_AVG_RT (Stroop Pre Congruent Average Response Time) | 28 | 0.57 | 0.95 | 0.7351 | 0.09911 |
| Stroop_POST_CON_AVG_RT | 27 | 0.51 | 1.01 | 0.6981 | 0.13540 |
| Stroop_PRE_CON_Score (Stroop Pre Congruent Score) | 28 | 89.58 | 100.00 | 96.5032 | 3.00376 |
| Stroop_POST_CON_Score | 27 | 81.25 | 100.00 | 95.7570 | 4.34264 |

| Experimental | N Statistic | Minimum Statistic | Maximum Statistic | Mean Statistic | Std. Deviation Statistic |
|---|---|---|---|---|---|
| DCCS_PRE | 27 | 71 | 146 | 113.30 | 22.250 |
| DCCS_POST | 27 | 68 | 146 | 116.22 | 23.919 |
| Flanker_PRE | 27 | 69 | 146 | 105.33 | 19.401 |
| Flanker_POST | 27 | 76 | 146 | 114.15 | 21.886 |
| Wordlist_PRE | 27 | 69 | 146 | 104.63 | 15.297 |
| Wordlist_POST | 27 | 80 | 139 | 105.37 | 15.272 |
| PVT_PRE | 27 | 83 | 126 | 96.93 | 11.645 |
| PVT_POST | 27 | 83 | 135 | 101.56 | 13.915 |
| GEC_PRE | 27 | 31 | 79 | 52.26 | 12.808 |
| GEC_POST | 25 | 32 | 84 | 58.16 | 15.929 |
| Stroop_PRE_AVG_RT | 27 | 0.50 | 1.03 | 0.7968 | 0.13481 |
| Stroop_POST_AVG_RT | 25 | 0.49 | 1.07 | 0.7230 | 0.14575 |
| Stroop_PRE_TOT_Score | 27 | 66.67 | 98.61 | 87.0889 | 8.99849 |
| Stroop_POST_TOT_Score | 25 | 66.67 | 98.61 | 87.8344 | 8.86915 |
| Stroop_PRE_INC_AVG_RT | 27 | 0.49 | 1.21 | 0.9067 | 0.17688 |
| Stroop_POST_INC_AVG_RT | 25 | 0.53 | 1.29 | 0.8234 | 0.18877 |
| Stroop_PRE_INC_Score | 27 | 20.83 | 100.00 | 74.2286 | 20.34886 |

| | | | | | |
|---|---|---|---|---|---|
| Stroop_POST_INC_Score | 25 | 37.50 | 100.00 | 75.5000 | 17.02381 |
| Stroop_PRE_CON_AVG_RT | 27 | 0.51 | 0.98 | 0.7418 | 0.11902 |
| Stroop_POST_CON_AVG_RT | 25 | 0.48 | 0.96 | 0.6730 | 0.12872 |
| Stroop_PRE_CON_Score | 27 | 77.08 | 100.00 | 93.8274 | 5.76377 |
| Stroop_POST_CON_Score | 25 | 81.25 | 100.00 | 93.6676 | 6.49133 |

When reviewing the test results conducted both pre and post (22 variables), separated by treatment (44 variables), 45.4% (20 variables) were within the approximately symmetrical range (-0.5 to 0.5), 43.2% (19 variables) were moderately skewed (-1 to 1) and 5 were highly skewed (greater than -1 or 1). The kurtosis values were more within spec with 77.3% falling within the normal range (within -1 to 1, for the SPSS output). Only 6% (3 variables) were somewhat leptokurtik (values greater than -1) and the remaining 16% were platykurtik with values greater than +1. While most of these leptokurtik and platykurtik variables were still between -1.5 to 1.5, two variables exceeded values of 3.00 (Stroop Post Incongruent Score, 12.358 and the Stroop Post Congruent Score, 3.345), indicating severely skewed tails.

Given the apparent skew in the data, the Shapiro-Wilk and Levene's test of normality were run and analyzed to determine the use of parametric and non-parametric tests. Two variables failed both tests (Stroop_PRE_CON_Score & Stroop_POST_CON_Score), an additional five tests failed the Shapiro-Wilk test (DCCS-POST, Wordlist-PRE, PVT-PRE, Stroop_PRE_INC_Score, Stroop_POST_INC_Score) and two additional variables failed Levene's test (Stroop_PRE_TOT_Score, Stroop_POST_TOT_Score). If either the pre- or post-variable failed a test of normality, for comparison purposes the entire variable was calculated using non-parametric tests. Table 5 illustrates the split between which variables were analyzed using parametric versus nonparametric tests.

**Table 5**

*Summary of Parametric versus Nonparametric Variables*

| Parametric Tests Used | Nonparametric Tests Used |
|---|---|
| Flanker | DCCS |
| GEC | Wordlist |
| Stroop_AVG_RT (Stroop Pre Average Response Time) | PVT |
| Stroop_INC_AVG_RT (Stroop Incongruent Average Response Time) | Stroop_TOT_Score (Stroop Total Score) |
| Stroop_CON_AVG_RT (Stroop Congruent Average Response Time) | Stroop_INC_Score (Stroop Incongruent Score) |
| | Stroop_CON_Score (Stroop Congruent Score) |

### ***Comparing Means***

In addition to reviewing skew and kurtosis, the T1 mean scores were compared for both the experimental and control groups to ensure limited variation prior to analyzing between-group differences (see Table 6).

**Table 6**

*Control versus Experimental T1 Mean Score*

| Tests | Control Mean | Experimental Mean | Difference* |
|---|---|---|---|
| DCCS | 109.62 | 113.30 | 3.68 |
| Flanker | 104.93 | 105.33 | 0.4 |
| Wordlist | 101.79 | 104.63 | 2.84 |
| PVT | 99.93 | 96.93 | 3 |
| GEC | 52.68 | 52.26 | 0.42 |
| Stroop_PRE_AVG_RT | 0.7933 | 0.7968 | 0.0035 |
| Stroop_PRE_TOT_Score | 90.7250 | 87.0889 | 3.6361 |
| Stroop_PRE_INC_AVG_RT | 0.9089 | 0.9067 | 0.0022 |
| Stroop_PRE_INC_Score | 79.1671 | 74.2286 | 4.9385 |
| Stroop_PRE_CON_AVG_RT | 0.7351 | 0.7418 | 0.0067 |
| Stroop_PRE_CON_Score | 96.5032 | 93.8274 | 2.6758 |

*Note*. *Differences are represented in absolute value

Following this, means were graphically compared both between groups (see Figures 12 and 13) and between genders for the experimental group (see Figures 14 and 15). While statistical analysis found limited support for a gender difference, comparing the means provides some support in some variation between male and female performance in the experimental group. Interestingly, when comparing the mean scores for the control group, it was actually the males that outperformed the females, which further supports the theory that there might be a trend in the experimental group females.

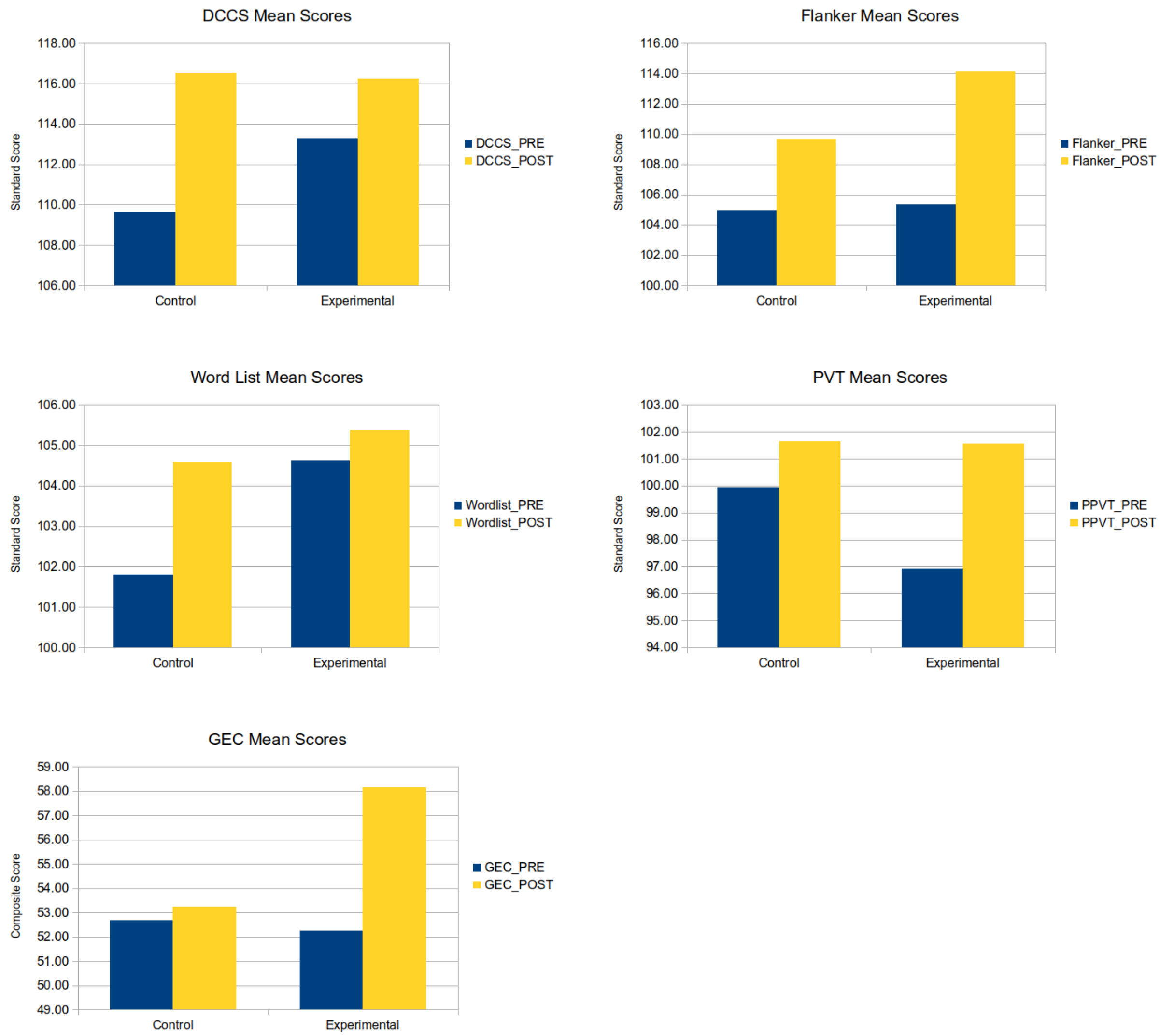


Figure 12: Pre/Post Scores, Less Stroop, by Treatment

*Note.* All tests except the GEC are displayed as age-corrected standard scores (mean of 100), while the GEC is reported as a t-score (mean of 50).

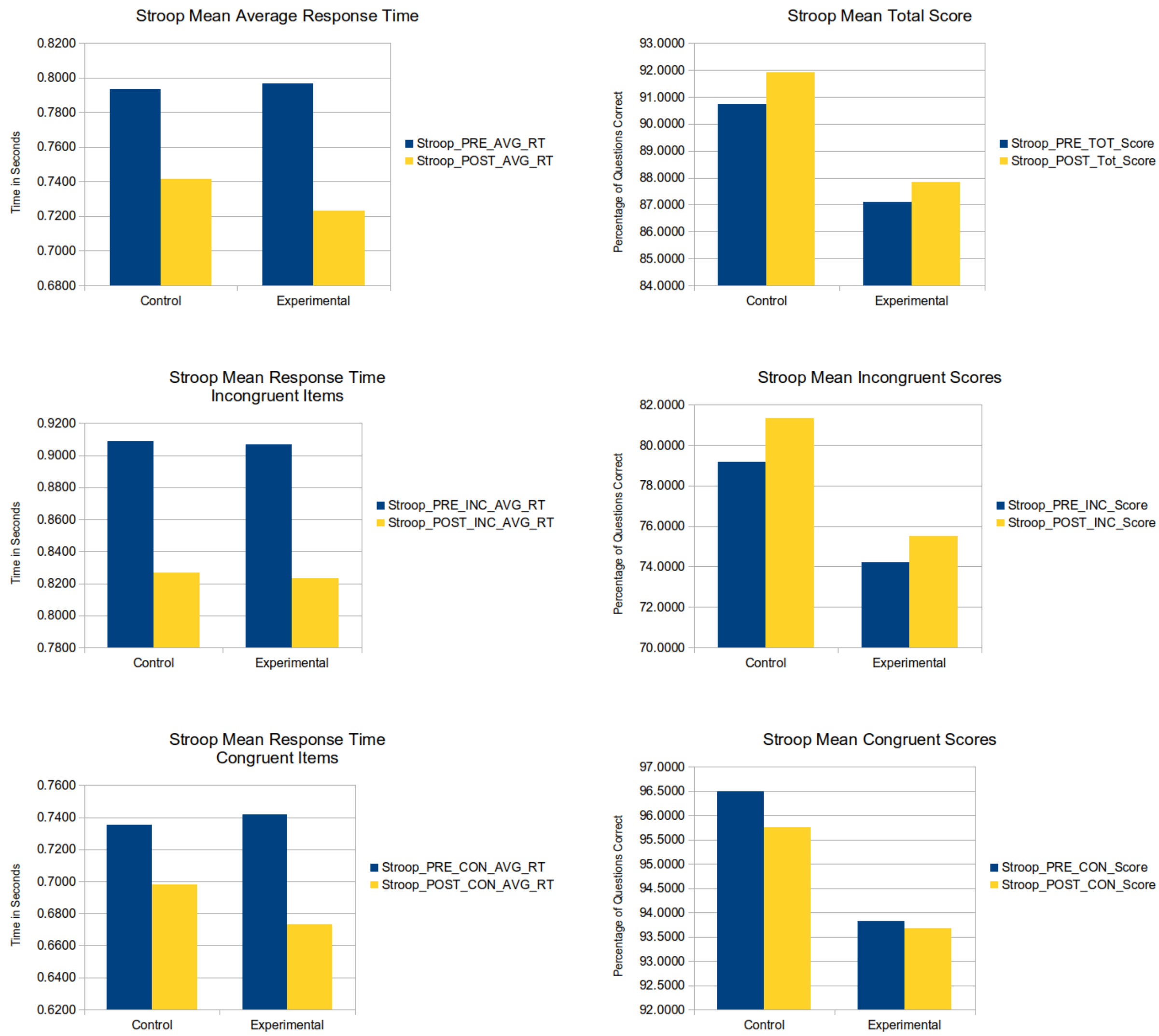


Figure 13: Stroop Pre/Post RT and Scores, by Treatment

*Note*. A shorter response time indicates a quicker response (left) while a higher percentage indicates a better score (right).

When comparing the means of the pre- and post-test scores for males and females within the experimental group only, females performed slightly better than the males on 8 of 11 measures (see Table 7). The time*gender results for these variables are listed in Table 10, and while these results were not statistically significant (p-values ranging from 0.198-0.856 for $n$=12 females), they are interesting and worth further exploration.

**Table 7**

*Comparison of Experimental Male and Female Mean Scores*

| | Male Mean Scores | | | Female Mean Scores | | |
|---|---|---|---|---|---|---|
| | Pre | Post | Difference | Pre | Post | Difference |
| DCCS | 113.40 | 114.27 | 0.87 | 113.17 | 118.67 | **5.50** |
| Flanker | 110.00 | 113.33 | 3.33 | 99.50 | 115.17 | **15.67** |
| Word List | 101.80 | 104.4 | **2.60** | 108.17 | 106.58 | -1.59 |
| PVT | 98.00 | 102.47 | 4.47 | 95.58 | 100.42 | **4.84** |
| GEC | 52.33 | 58.77 | 6.44 | 52.17 | 57.50 | 5.33 |
| Stroop Avg RT* | 0.799 | 0.739 | -0.060 | 0.794 | 0.703 | **-0.091** |
| Stroop Inc RT* | 0.916 | 0.846 | -0.070 | 0.895 | 0.795 | **-0.100** |
| Stroop Cong RT* | 0.742 | 0.686 | -0.056 | 0.742 | 0.657 | **-0.085** |
| Stroop Total Score | 86.02 | 86.71 | 0.69 | 88.43 | 89.27 | **0.84** |
| Stroop Inc Score | 73.61 | 72.62 | -0.99 | 75.00 | 79.17 | **4.17** |
| Stroop Con Score | 92.78 | 93.16 | **0.38** | 95.14 | 94.32 | -0.82 |

*Note.* While technically the females also outperformed the males on the GEC with a lower score, given that both scores increased where they should have decreased, neither gender is noted as outperforming the other in this measure.
The bolded score indicates the better score (male or female).
*RT is measured in seconds and is given an extra decimal point given the brevity of the measure.

While not all the tests show great disparity in their differences (for example, male versus female PVT), others show more improvement, such as the females' performance in the DCCS and Flanker tasks. Overall, it is interesting to note the differences and potential trends in the post-score improvements. These differences are further illustrated by the graphs in Figure 14 and 15 below.

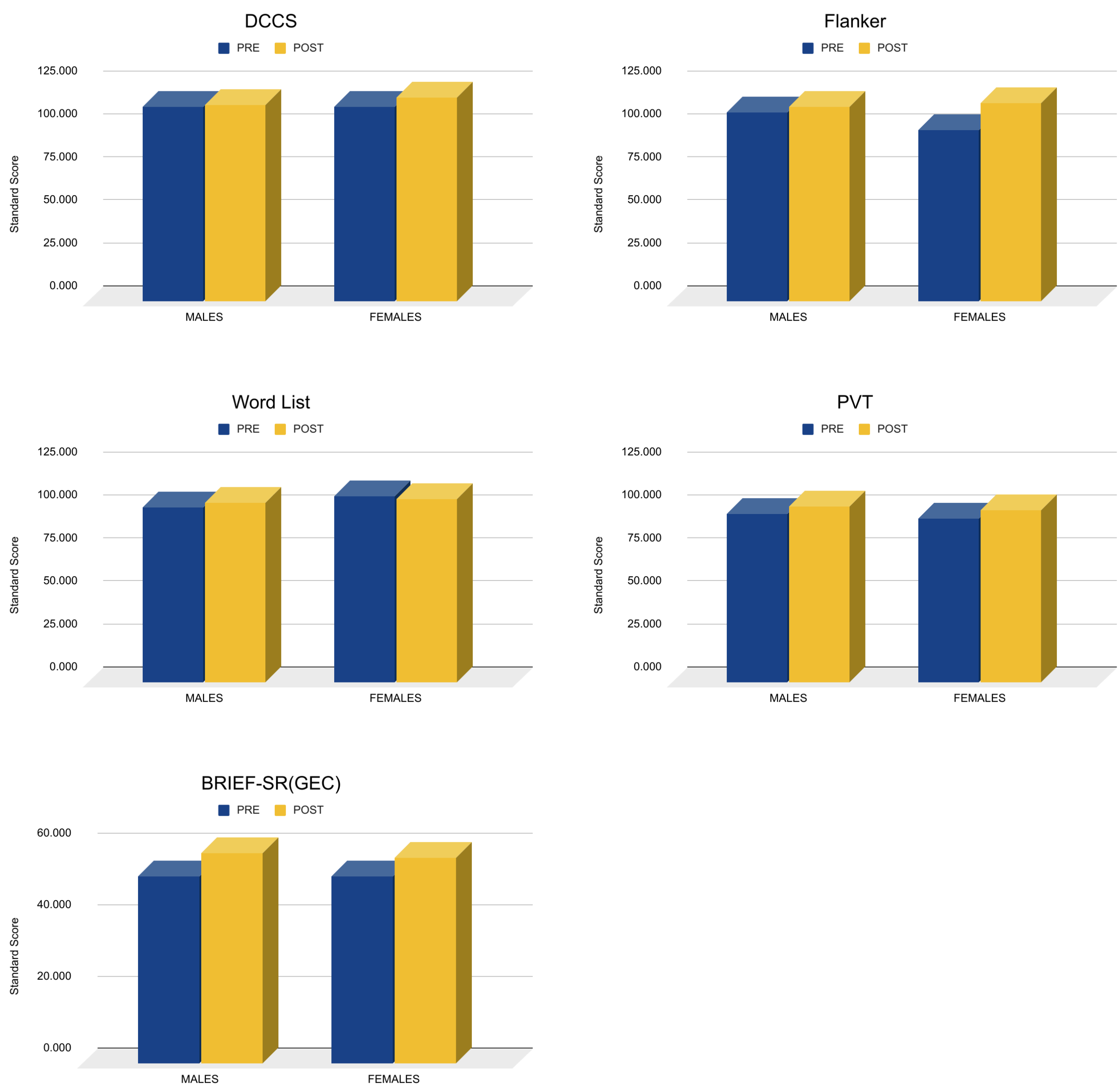


Figure 14: Experimental Group Pre/Post scores, Less Stroop, by Gender

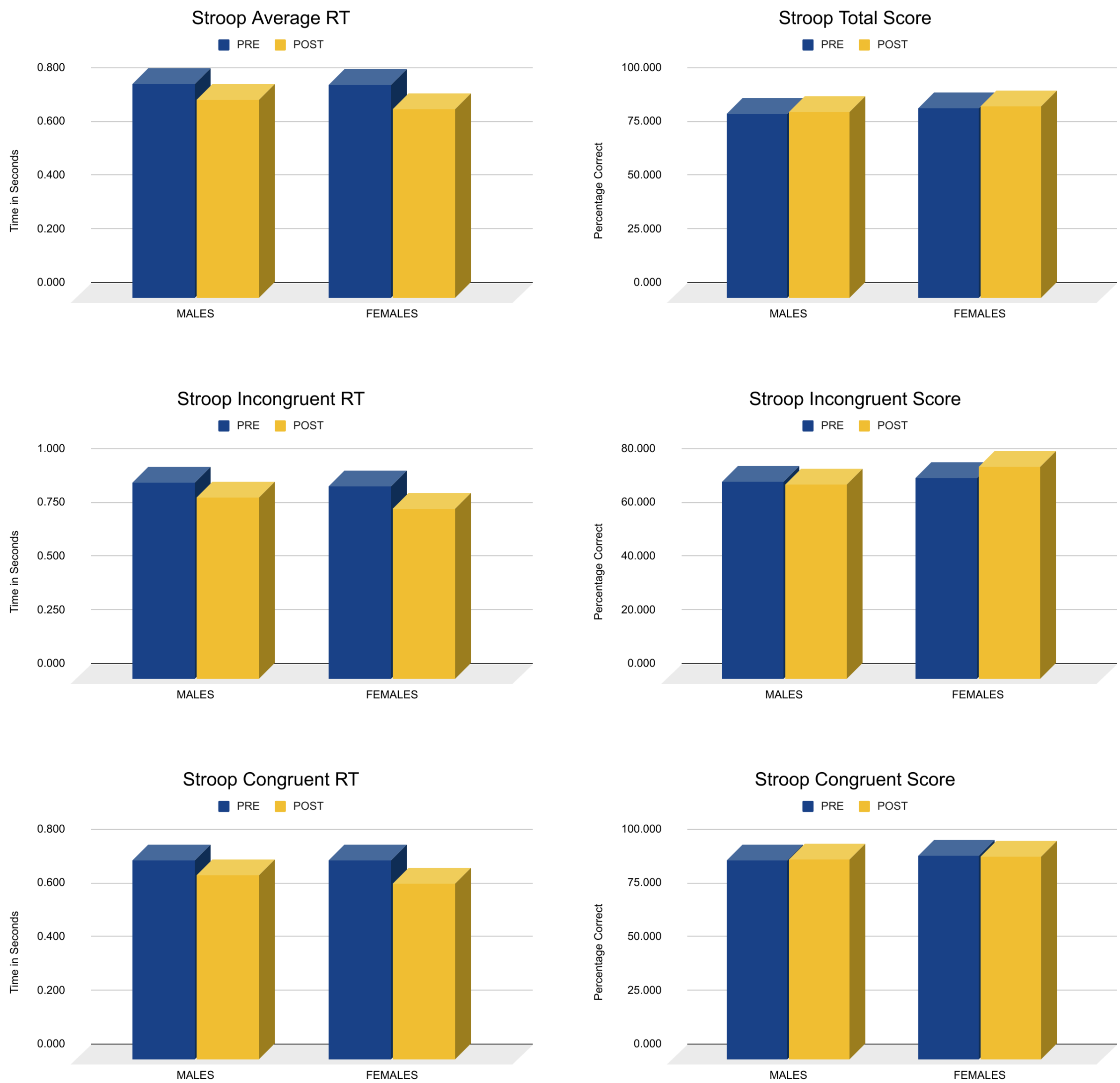


Figure 15: Experimental Group Pre/Post Stroop RT and Scores, by Gender

*Note*. A shorter response time (left) indicates a quicker response while a higher percentage indicates a better score (right).

### *Z-Scores & Outliers*

Prior to moving into further tests, the researcher also transformed the dependent variables into z-scores and inspected the data for any outliers which may be contributing to the skew or kurtosis of the data. During this inspection it was found that three z-scores fell outside of the 3.0 outlier cutoff; all three scores belonged to the same participant (GHM12, control group) and all three scores were related to the Stroop pre-test. As a result of these findings, the conclusion was drawn that there was a level of inaccuracy during the Stroop pre-test, and all Stroop pre-scores for GHM12 were removed prior to further analyses.

Having discovered these outliers, the researcher further investigated the data for additional outliers using both box plots and interquartile range (IQR) in SPSS. Given the shape of the data, only seven variables required any adjustment (see Table 8 below). Applying the formula for IQR identified more outliers than using the box plots (23 versus 14), and as most of the box plot outliers were also identified by IQR (11 of 14), and this data set was already limited, the researcher chose to only remove the outliers identified by the box plots for a total of 14 removals. It is worth noting that the researcher reanalyzed the descriptive statistics after adjusting for the outliers but the shape of the data did not change significantly and the normality of the data remained the same as was noted above in Table 5.

**Table 8**

*Box Plot Outliers with IQR for Reference*

| | IQR (Lower-Upperbound) | Participant & Score |
|---|---|---|
| Wordlist Pre | 75.5 - 127.5 | COF12 (146)<br>DKM12 (69) |
| Wordlist Post | 65 - 145 | EFM13 (146) |
| PVT Pre | 71.5 - 123.5 | EBM12 (135)<br>ROM12 (126)<br>GHM12 (126)<br>CPM12 (74)** |

| | | |
|---|---|---|
| Stroop Post Total Score | 75.005 - 108.325 | GHM12 (76.39)*<br>KLF13 (69.44)<br>LMM12 (66.67) |
| Stroop Post Inc RT | 0.3265 - 1.3065 | LBM12 (1.31)<br>ASM12 (1.29) |
| Stroop Pre Con Score | 84.375 - 109.375 | KSF12 (77.08) |
| Stroop Post Con Score | 79.175 - 112.495 | GHM12 (81.25)* |

*Note*.
*GHM12's scores, while identified as an outlier with the box plots, were within the specified IQR, however, given the previous outliers identified with this case the researcher removed these scores
**Identified by box plots only

## Paired-Sample T-Tests & Non-Parametric Equivalents

### ***Paired-Sample T-Tests***

Paired-Samples T-Tests were run for the parametric variables listed above in Table 5 to analyze whether or not there were differences between pre- and post-testing, and the output was split by case (control versus experimental). It is important to note that for the study's hypothesis to be true (that the curriculum of games improves EF) that the scores of certain variables must increase, and others must decrease. For example, the response time variables of the Stroop Test should decrease if participants' EF improves, while their Stroop *scores* should increase. Similarly, while the standard scores of the DCCS, Flanker and Wordlist should increase with EF improvement, the GEC as calculated from the BRIEF-SR should *decrease.* Due to the directional nature of these scores, both the one-tailed and two-tailed results for the control and experimental groups are displayed in Table 9.

**Table 9**

*One-Tailed & Two-Tailed T-Test Results, Split by Treatment*

| Control Results | | |
|---|---|---|
| | One-Tailed | Two-Tailed |
| Flanker | 0.056 | 0.112 |
| GEC | 0.325 | 0.650 |
| Stroop_avg_rt | 0.005 | 0.009 |
| Stroop_inc_avg_rt | <0.001 | 0.001 |
| Stroop_con_avg_rt | 0.020 | 0.039 |
| **Experimental Results** | | |
| | One-Tailed | Two-Tailed |
| Flanker | 0.037 | 0.073 |
| GEC | <0.001 | <0.001 |
| Stroop_avg_rt | <0.001 | 0.001 |
| Stroop_inc_avg_rt | 0.002 | 0.005 |
| Stroop_con_avg_rt | <0.001 | 0.001 |

While we see significant results ($p$ = 0.05 and less) in many of the tests in both groups, it is worth noting that the results for the experimental group are quite a bit stronger. This suggests that while there were significant improvements in scores between pre- and post-testing, perhaps we are also seeing a practice effect due to the limited time between T1 and T2 for the control group. The experimental group completed T1 and T2 an average of 12.5 weeks apart, whereas the control group completed T1 and T2 an average of 6 weeks apart.

Finally, one important finding worth further exploration is the GEC. As previously mentioned, for the GEC to support the hypothesis, it would need to decrease, however, upon closer inspection of the scores, the GEC scores actually went up in post-testing. As a T-Score, the GEC has a mean 50. The control group saw an average pre/post increase of only 0.56 points (52.68 to 53.24), whereas the experimental group saw an average increase of 5.9 points (52.26 to 58.16). Further parsing the data, within the control group, females saw an increase of 4.3, while males saw a decrease of 1.44. Within the experimental group, females saw an increase of 5.33 and males saw an increase of 6.44. While these increases actively dispute the hypothesis, these results, and their potential implications, will be explored further in the discussion section.

***Wilcoxon Signed-Rank & Mann-Whitney U Tests***

Since half of the variables did not meet the tests of normality, non-parametric tests were used to investigate their significance. The Wilcoxon Signed-Rank test was used to compare the pre- and post-test scores within subjects and split by treatment, whereas the Mann-Whitney U test was used to compare the differences in change between the control group and the experimental group.

When running the Wilcoxon Signed-Rank test with the control group data, no significant effects were found. However, when running this test with the experimental group, the PVT was significant at $p = 0.032$. The Mann Whitney U test reported no significant results. For good measure, the Friedman test was also run to evaluate the effect of the repeated measures with the subjects and split by treatment. Similarly to the Wilcoxon Signed-Rank test, the Friedman test found no significant effects within the control group but found a significant result for the PVT within the experimental group ($p = 0.046$).

**Parametric Multivariate & Repeated Measures Tests**

Parametric testing continued for the five variables that met the conditions of normality (see Table 5). With multiple independent and dependent variables within the study, the researcher analyzed a number of multivariate effects to determine any interrelationships; MANOVAs conducted included dosage, gender, treatment and gender*treatment. The only interaction that was approaching any significance in these tests was gender and treatment ($p = 0.151$).

When running the repeated measures tests on each of the parametric variables, significant results were found for time (pre-post of dependent variables) and one significant result was found for the time*treatment interaction (GEC), these results are detailed in Table 10:

**Table 10**

*Repeated Measures Results for Parametric Variables*

| | Flanker | GEC | Stroop_avg_rt | Stroop_inc_avg_rt | Stroop_con_avg_rt |
|---|---|---|---|---|---|
| Time | **0.018** | **<0.001** | **<0.001** | **<0.001** | **<0.001** |
| Time*Treat | 0.329 | **0.033** | 0.302 | 0.735 | 0.148 |
| Time*Gender | 0.482 | 0.198 | 0.856 | 0.701 | 0.822 |
| Time*Treat* Gender | 0.137 | 0.105 | 0.461 | 0.432 | 0.516 |

Given the consistency of significant findings for time, the researcher further explored the effect size and power of the significant findings. The partial eta squared denotes a medium effect size between 0.06-0.139, and a large effect size above 0.15. All results are above 0.06, with four of six results being above 0.139. Furthermore, medium power is typically indicated between 0.51-0.80, with high power being indicated by results greater than 0.80. In these results, two of the results have medium-observed power, with the remaining four having high-observed power. Two of the results have both medium effect and medium power, these being the Flanker (time), and the GEC (time*treatment). These supporting findings are detailed in Table 11.

**Table 11**

*Effect and Power of Significant Results*

| Significant Result | Partial Eta Squared | Observed Power |
|---|---|---|
| Time - Flanker (0.018) | 0.102 | 0.667 |
| Time - GEC (<0.001) | 0.219 | 0.952 |
| Time*Treatment - GEC (0.033) | 0.089 | 0.574 |
| Time - Stroop_avg_rt (<0.001) | 0.290 | 0.990 |
| Time - Stroop_inc_avg_rt (<0.001) | 0.303 | 0.991 |
| Time - Stroop_con_avg_rt (<0.001) | 0.261 | 0.979 |

**Bootstrapping**

During data analyses, the researcher applied bootstrapping where SPSS would allow to explore whether or not a more robust sample would provide more normalized or more significant results. In all cases the changes provided by bootstrapping were limited or insignificant, as can be seen in a comparison of the original versus bootstrapped descriptive statistics in Table 12.

**Table 12**

*Original versus Bootstrapped Descriptive Statistics*

| Test | Mean Statistic | Bootstrapped Mean | Std. Deviation Statistic | Bootstrapped Std. Deviation |
|---|---|---|---|---|
| DCCS_PRE | 111.39 | 113.59 | 22.758 | 22.605 |
| DCCS_POST | 116.38 | 118.62 | 22.706 | 22.421 |
| Flanker_PRE | 105.13 | 107.08 | 19.889 | 21.718 |
| Flanker_POST | 111.82 | 109.21 | 21.374 | 20.458 |
| Wordlist_PRE | 103.00 | 102.26 | 15.465 | 13.816 |
| Wordlist_POST | 104.22 | 104.56 | 15.372 | 15.245 |
| PVT_PRE | 97.19 | 98.28 | 10.213 | 10.473 |
| PVT_POST | 101.61 | 100.46 | 13.138 | 11.429 |
| GEC_PRE | 52.47 | 52.44 | 12.713 | 12.863 |

| | | | | |
|---|---|---|---|---|
| GEC_POST | 55.52 | 55.62 | 15.861 | 16.393 |
| Stroop_PRE_AVG_RT | 0.7950 | 0.8000 | 0.12325 | 0.12229 |
| Stroop_POST_AVG_RT | 0.7326 | 0.7334 | 0.14416 | 0.13286 |
| Stroop_PRE_TOT_Score | 88.9400 | 89.7082 | 7.81973 | 8.05243 |
| Stroop_POST_Tot_Score | 91.1290 | 91.4528 | 5.79492 | 5.96541 |
| Stroop_PRE_INC_AVG_RT | 0.9079 | 0.9082 | 0.16222 | 0.15720 |
| Stroop_POST_INC_AVG_RT | 0.8063 | 0.8212 | 0.15342 | 0.16120 |
| Stroop_PRE_INC_Score | 76.7428 | 78.0991 | 18.38567 | 18.55359 |
| Stroop_POST_INC_Score | 79.9681 | 82.5856 | 14.48154 | 12.23635 |
| Stroop_PRE_CON_AVG_RT | 0.7384 | 0.7457 | 0.10837 | 0.10987 |
| Stroop_POST_CON_AVG_RT | 0.6860 | 0.6891 | 0.13154 | 0.12360 |
| Stroop_PRE_CON_Score | 95.5250 | 95.7269 | 4.05561 | 4.37961 |
| Stroop_POST_CON_Score | 95.0173 | 95.6203 | 5.23908 | 4.93979 |

**Excluded Data**

While data was collected and compiled for the LBBT, the processing and cleaning of this data requires further input and instruction from another lab. As this method was only just published (Aguilar et al., 2021), the researcher requires additional clarification on scoring methodology in order to fully interpret the results. These results will be analyzed and published separately from this thesis document.

# CHAPTER 5: DISCUSSION & FUTURE DIRECTIONS

## Summary

Starting with the paired-samples t-tests, and the Wilcoxon signed-rank test, a number of significant findings are present in both the experimental and control groups as outlined in Table 13:

**Table 13**

*Tests with Significant Results by Parametric/Non-Parametric and Treatment*

| **Parametric Tests Used** | | | |
|---|---|---|---|
| **Control Group** | Result | **Experimental** | Result |
| Flanker | NS* | Flanker | 0.037 |
| GEC | NS | GEC | <0.001 |
| Stroop_AVG_RT | 0.005 | Stroop_AVG_RT | <0.001 |
| Stroop_INC_AVG_RT | <0.001 | Stroop_INC_AVG_RT | 0.002 |
| Stroop_CON_AVG_RT | 0.020 | Stroop_CON_AVG_RT | <0.001 |
| **Non-Parametric Tests Used** | | | |
| **Control** | Result | **Experimental** | Result |
| DCCS | NS | DCCS | NS |
| Wordlist | NS | Word List | NS |
| PVT | NS | PVT | 0.032 |
| Stroop_TOT_Score | NS | Stroop_TOT_Score | NS |
| Stroop_INC_Score | NS | Stroop_INC_Score | NS |
| Stroop_CON_Score | NS | Stroop_CON_Score | NS |

*Note*. *NS means not significant

Of 11 variables, the paired samples t-test and Wilcoxon signed rank test identified three instances of significance for the control group and six for the experimental group, with stronger results for the experimental group. As referenced in the results section however, a potential confound was found when collating the data; when reviewing the time between T1 and T2 for both treatments, it was found that the experimental group had an average of 12.5 weeks (88 days) between the instances of testing, whereas the control group only had an average of 6 weeks (44 days) between testings. This will be further explored in the limitations section, however, this may allow for a stronger practice effect for the control group, which may partially explain some of their more significant results.

For the repeated measures test and the Friedman test, most of the significant results are in the time condition with GEC also showing significance in the time*treatment condition ($p$ = 0.033; see Table 13 for other results). Based on the partial eta and the observed power of these time effects, the results are strong. For the time*treatment results, however, similar to the t-tests, it's possible that some of the results are being diminished through the potential increased practice effect in the control group.

## Interpretations and Implications

### *Non-Normality*

With half of the variables failing the tests of normality, it is likely that the sample was not quite large enough to achieve a standard distribution. With only $n$ = 56, and 22 females overall, further numbers may have better supported the sample. However, the researcher did attempt to employ bootstrapping with minimal changes in the shape of the data.

### *Comparing Means*

Despite limited normality, the mean scores provide a level of confidence in the results. When comparing the pre-tests of the two treatments, there is some variability across the two groups (in particular the Stroop incongruent score, the DCCS and the PVT; see Figures 12 and 13), but the variability was limited and went both ways as the control group performed better on certain tasks such as the PVT and Stroop scores. The pre-scores and demographics of the two treatment groups suggest that there were no remarkable differences between the two groups prior to the intervention (see Table 6).

When reviewing the differences within the experimental group post-intervention, there are also some potentially interesting findings related to gender. While these results do not come through as statistically significant, it is very possible that with only $n$ = 26 experimental participants (12 female), that there were not enough cases to provide sufficient power and effect size. That being said, while comparing the means of these two demographics (see Figures 14 and 15 and Table 7) it is worth noting that the females outperformed the males on 8 of 11 tasks. In some of these tasks the females only slightly outperformed the males, such as in the PVT where males saw an increase of 4.47 points, while females saw an increase of 4.84. However, in other tasks, such as the Flanker, females saw an increase of 15.67 points versus only 3.33 for the males. Similarly, for the DCCS, females saw an increase of 5.58 whereas the males increased by only 0.87 points.

As discussed in Chapter 2, and as discussed by Kolb et al. (2023), puberty and gonadal hormones have an impact on adolescent brain plasticity. As females tend to reach puberty in advance of males, the females are potentially seeing greater benefits from the intervention due to the intervention being delivered during the critical period when their brains are most plastic.

Conversely, it is possible that the males have not fully entered puberty and are seeing less growth as they are not fully in that critical window. While no demographic or biological data was collected to confirm participants' stage of puberty, and thus we cannot conclusively correlate the two, this potential should be explored further to better understand not only the impact of EF interventions, but to better understand this critical period of adolescent brain development.

***T-Tests***

The paired samples t-tests provide the initial support towards rejecting the null hypothesis in favour of the theory that there is in fact a change in participants post-intervention. While both the control and experimental groups see significant results, the fact that the experimental group's results are stronger suggests that there is more than a practice effect, which may be the case for the control group.

***GEC scores***

As previously mentioned, for the BRIEF GEC score to indicate improvement, it would need to decrease. However, the experimental group saw a significant *increase* in their scores post intervention (11.3%), while the control group saw a negligible increase (1.06%). In the experimental group, both male (m) and female (f) pre-GEC scores were comparable (f = 52.17, m = 52.33) and males increased slightly more than the females in their post-scores (f = 57.5, m = 58.77). An increase in scores would generally indicate a worsening in the GEC score, however, given the decreases or relative stability in the remainder of the scores, it would be unlikely that the increase would indicate an actual decrease in EF. An alternative hypothesis is that participants in the experimental group had the opportunity to reflect on their own capabilities during the intervention and upon completing the post-BRIEF-SR, more accurately rated their skills.

### *Repeated Measures*

With the parametric multivariate and repeated measures tests we see significant results for all variables in the time factor and one significant result in the time*treatment factor (GEC). This tells us that the changes between T1 and T2 are statistically significant, as is confirmed by the power and effect size. This could potentially suggest that the changes we are seeing are a result of the intervention. However, moderation is needed as the time*treatment results are less prominent with only one variable reaching significance threshold. As previously alluded to, with the control group having completed T2 in half the time of the experimental group, there is a strong possibility that we are observing an attenuation of the time*treatment effects as a result.

### *Tests and Scores*

While the non-parametric tests may not hold as much strength or weight as the parametric tests, when we review all the tests with significant results throughout the study, we are left with the Flanker,  the BRIEF-SR, the Stroop, and the PVT.

The Flanker task, developed by Eriksen and Eriksen (1974), is designed to assess inhibitory control, among other variables such as attention (Baghdadi et al., 2021; Eriksen & Eriksen, 1974). Likewise, the Stroop task also assesses inhibitory control (Basu, 2022; Jensen & Rohwer, 1966). While the DCCS also requires some inhibitory control (to suppress one rule in favour of another), more so, this task relies heavily on task switching, or cognitive flexibility (Doebel & Zelazo, 2015). It is then possible, that in assessing for multiple EFs, that over the course of this intervention, only inhibitory control has been noticeably impacted as these are the tests of EF that see significant results.

## Limitations

While there are many promising findings and areas for potential future analysis, there are limitations that may have had effects on the data, these include testing in the school environment, the non-randomized nature of the classrooms, and the time difference between treatment groups and T1 and T2.

### *The School Environment*

Conducting testing in the school environment has both its benefits and drawbacks. While students are more likely to be comfortable in their normal daily environment, the school environment can be busy, noisy and occasionally unpredictable. While there were very few unanticipated schedule changes, we were only able to conduct testing during certain school periods and in certain locations. This occasionally presented challenges when attempting to coordinate testing. In addition to this, testing in the school environment was sometimes loud and disruptive; there were a handful of instances where participants' tests were interrupted or disrupted due to noise or commotion in the hall adjacent to the testing room. None of these instances were significant enough to warrant restarting a test, but there were distractions that would have minor impacts on scores.

### *Non-Randomized Nature of the Classrooms*

For the recruitment process, teachers self-selected if they were interested in trialing the intervention. In addition to this, as is normal in schools, class lists are generated at the beginning of the school year with a variety of factors in mind. As the control and experimental group both self-selected, and their class lists were already established, the study could only be quasi-experimental given these conditions. In addition to this, the teachers expressed that both classrooms were quite a bit different with respect to behaviors and academics. Although classes were generally matched in size and composition (control: *n* = 29, f = 10; experimental: *n* = 27, f= 12), the behavioural differences between the two groups could have potentially contributed to the skew in the shape of the data.

***Time Difference Between Treatments for T1 and T2***

Pre-testing for the experimental group began at the end of January and concluded on the 10th of February 2023. Pre-testing for the control group began at the beginning of March and concluded March 15th, 2023. The learning curve for testing was steep at the beginning as the researcher was training multiple students to assist with the testing. The intervention was designed to be used over 6 weeks, and given school breaks and in-service days, the experimental group started on the week of March 6th, and concluded the intervention on April 24th. As we were approaching the end of the school year and final exams, and with testing taking approximately 45 minutes per participant, the researcher made the decision to begin post-testing with the control group while the experimental group finished the intervention. The time between T1 and T2 was not calculated as all the testing was conducted back-to-back and a discrepancy was not factored in given the importance of wrapping up testing. Ultimately, in review, given the learning curve in testing during T1 and the amount of time needed to complete the intervention, the gap between T1 and T2 for the experimental group was considerably longer than the control. As mentioned previously, testing in a school setting is a dynamic process, and while the decision to test the control group prior to the experimental group for T2 was made out of consideration for the school timeline, this may have ultimately impacted data collection.

***Data Limitations***

Finally, with respect to data collection and analysis, it is noted that an analysis of the LBBT may add another angle to the results. This data is still being carefully reviewed and scored, with further oversight needed prior to analysis. Once a detailed review of the data and methodology is completed, further analyses will be conducted to review both the LBBT data, and the LBBT data in conjunction with the other results.

**Future Directions**

Play is an important process, even for adolescents, and has the potential to positively impact EF. Play, whether collaborative or competitive, provides youth with an opportunity to develop social skills, practice theory of mind and turn-taking, and to understand rules and boundaries (Else, 2009; Lillard, 2015). In addition to this, as referenced in Chapter 2, both social and solo-play can also have varying educational impacts (Kaye & Bryce, 2014; Ruipérez-Valiente & Kim, 2020). While play behaviour in adolescents is notoriously understudied, the benefits of play across the ages are undeniable. These benefits include modifying executive functions and training self-regulation (Shaheen, 2014), honing sensory systems and social abilities (Yogman et al., 2018), providing opportunities for parents to connect with their children (Ginsburg et al., 2007), and these benefits persist down to the molecular and cellular levels of the brain (Yogman et al., 2018).

While research is limited in many areas of adolescent development, research has shown that play has been on the decline in elementary schools (Ginsburg et al., 2007). Generally, once children move out of elementary school and into middle or high school, recess is no longer available to them; nor is there as much access to play “materials.” The school day for adolescent children generally consists of two short breaks–one in the morning and one in the afternoon–and a lunch period. While these breaks allow for some socialization, they are also often co-opted for academic catchup, tutoring, and often for much needed bio-breaks. In addition to this, yet outside the scope of this paper, is also the issue of technology acting as a barrier to socialization and play during the school day. Therefore, if we know that play is important, and we know that adolescents–in particular–don't get enough of it, the intention of this study is then to bring play into the school day, and into the classroom, in a way that still supports cognitive and academic growth. The curriculum presented herein supports adolescent EF through teacher-led play throughout the school week.

While instructional time in schools is always a prized commodity, the BBT adolescent curriculum looks to support EFs which are critical to not only academic success, but also success throughout life. This curriculum seeks to support teachers and caregivers in building crucial skills in their students, such as inhibitory control, cognitive flexibility and working memory, so that they can maximize their time by teaching regulated, flexible and plastic brains. In order to support a better understanding of adolescent EF, and the impact of interventions such as the one presented in the current study, further research is needed. By replicating this study considering the aforementioned limitations, as well as building onto the existing methodology, the current findings could be supported while still building onto them. Finally, elaboration of the current study would aid in consolidating some of the existing findings related to EF, as well as to better establish the literature with respect to supporting adolescent learners.

***Literature***

As identified in the literature review, while there is a breadth of literature explaining EF in general, there is limited research on EF with relation to adolescents. Further exploration of this intersection is essential for a better understanding of the development of EF in adolescence which is a critical developmental period. With the understanding that puberty and adolescence is a tumultuous time—as illustrated by the emotional face-matching task described by Blakemore (2008) in Chapter 2—further research in this area is essential to fill the gap in the literature.

One large gap in the literature that needs further exploration is the concept of adolescent play. By this we do not mean studies solely investigating playing video or computer games, but studies that investigate the form and function of adolescent play, including but not limited to digital media. Of particular interest in our current digital world, would be a collection and comparison across multiple generations that investigates the predominant trends in adolescent play. We know that our lives are more digital than ever, but studies involving, and comparing, in-person adolescent play could also provide important insights into aspects of mental health, academics and life success.

Furthermore, adolescent play may still need its own definition. Many current definitions of play (see Else, 2009; Lillard, 2015) posit that play should be spontaneous and independently instigated. However, when we look at our education systems, there are many teacher-driven activities that would otherwise meet the definitions of "play" with the exclusion of these criteria. The definition provided by the author herein is a good start, however a separate definition is needed to address adolescent solo-play, a topic not covered in this paper. In addition to this, there is still some debate as to whether team sports constitute play, and due to adolescents' increasing concerns regarding self-esteem and self-image, there are further questions regarding whether physically active games may constitute play or induce anxiety for teens.

Despite our advances in knowledge and technology, much remains to be understood about the human brain, and in particular the rapid and complex changes that are occurring during adolescence. While the landscape of teenage play has shifted dramatically since the introduction of video and computer games, play—and in particular in-person social play—is still a developmentally important activity throughout a persons' lifespan.

*Methodology*

While the methodology of the current study is sound and provided a great deal of value, there are aspects of the study that can be improved for future investigations. This could include further research into the curriculum, expanding the research to multiple schools, assessing the impact of the curriculum on different ages, as well as including an assessment that looks at changes in adolescents' emotional regulation capabilities.

**Curriculum Exploration.** The current curriculum, conceptualized by two undergraduate students in the Gibb lab (Fleischman et al., 2022) was piloted in the summer of 2022 with the final curriculum coming together for fall 2022. With the initial success of the program, future exploration could further support and build the curriculum. For example, as discussed in the literature review, both the physical activity games and tabletop games potentially have different impacts on participants' EF. Therefore, it would be worthwhile to explore the potentially different effects between the two by separating them into different experimental groups and testing for any differences. Furthermore, it could be beneficial to explore additional games that may also add value to the current curriculum.

**Multiple Sites.** The current study worked with one local middle school which kindly provided both a control and experimental group. As previously referenced, with $n = 56$, more participants could potentially provide more robust results. While this could be accomplished in one school, expanding to multiple schools could provide further substantiation of positive results by showing improvement in unrelated experimental sites.

**Impact of Age.** As referenced earlier, there is the potential that the female experimental participants saw increased improvement over the males. It would be prudent to conduct this study across different grades, ages and stages of puberty, to determine at what age(s) participants see the most improvements. Further exploration of this variable could indicate where additional study is warranted to identify the impacts of gonadal hormones on EF improvement.

**Emotional Regulation.** In addition to the skills of working memory, cognitive flexibility and inhibitory control, additional measurements could be added to the study to identify whether there would be improvements to participants' emotional regulation. This could be added through additional validated instruments as well as by collecting school data to indicate if the number of student infractions (ie. detentions, suspensions, etc.) decrease. This addition would be an important step in the research, as emotional regulation abilities are the other side of the coin to the EF measures herein and are equally important for adolescents.

***Replication & Expansion***

One proposed method to better understand the development of EF is to not only replicate the current study but expand it by adding additional components such as hormonal testing. One trend from the current study suggests that the females in the experimental group saw greater improvement than their male counterparts. Given that the current study tested Grade 7 students (ages 12-13), and given that females enter puberty earlier than males, it is possible that we have identified a key period for female participants. Given this potential outcome, including a method of hormonal testing to the current study would provide another level of exploration and potential confirmation for the effects seen in the current study.

Furthermore, there is potential to more completely understand the impact of EF interventions by looking into employing fMRI technologies. While the human connectome project is now working on exploring youth ages 5-21 (Somerville, 2018), a detailed exploration of the neurobiological changes pre- and post-intervention would help further establish adolescence as a critical period during which we can support neurobiological development. This could be done by choosing psychometric tests that are feasible to administer within the fMRI in conjunction with control scans.

In general, there is also potential to see more conclusive results with different psychometric tests. For example, while the Word List test was a complex and detailed task that required participants to actively engage their working memory (Tulsky et al., 2014), there are some aspects to this task that, upon reflection, may have posed challenges for participants. Firstly, because this task is word-based (versus digit-based, such as in the backwards digit span task), there is potential that participants whose first language is not English may have been inadvertently disadvantaged by this task. While the same could be said for the backwards digit span, numerals are more universal with more logical rules and applications. Furthermore, there is some evidence to suggest that the Word List task may not be as accurate in youth with ADHD (Jusko et al., 2021) which may have implications for testing within the school and general population. Future analysis and publication could include a thorough cross-analysis of all disclosed factors (for example primary language and medical conditions) with respect to individual tests. This, combined with the literature, could potentially provide further insight into the validity of these instruments. Therefore, while replication is an important part of future research, once this is completed, it may be worthwhile to further explore which tasks may best assess diverse adolescent populations.

Finally, as it is known that relationships are one of the most powerful influencers of student achievement (Sointu et al., 2017; Konishi et al., 2010), further exploration into the impact of a play-based intervention on teacher efficacy is warranted. Leveraging existing scales, a future study could include a measure of teacher efficacy or relationship. Should no scale exist to properly assess this combination of variables, the creation and validation of a scale that helps to assess the impact of an improved classroom relationship on teacher efficacy could be beneficial.

**Conclusion**

While the results of the current study are promising, there still remains much to explore with respect to adolescent EF. This thesis explores whether or not a selection of games, played a minimum of three times a week in the classroom, will improve EF. The hypothesis was that the purposeful inclusion of games, chosen to build executive functions, would enhance the EF skills of inhibition, working memory and cognitive flexibility in the participants (Sala & Gobet, 2016). While the findings discussed in the results and discussion section of this document are in favour of the hypothesis, caution should be applied in interpreting the results; they are favourable, but they are also modest and must be interpreted as such. In addition to this, the hypothesis states that we will see improvements in all three listed EF, and the results do not support significant improvement in all areas.

For example, when considering the tests that provided significant results (Flanker, Stroop response times and PVT), there is support for the finding that participants saw improvements in inhibitory control. The BRIEF-SR generates the GEC, which is a composite score of eight different subscales, therefore providing an overall score of EF. While the GEC findings of this study are novel, they cannot be used in support of the hypothesis given they trend in the wrong direction. Furthermore, the DCCS and Word List, tests that would show changes in cognitive flexibility and working memory, did not reflect any significant changes in the current study. Therefore, while the results provide a level of support for the hypothesis, it can not be said that the hypothesis has been confirmed in full.

When reflecting on the methodological processes used to answer this hypothesis, much can be said in favour of the procedures, and much can still be learned and improved upon. As mentioned throughout the last two chapters, there are tests and processes that may benefit from revision, but ultimately, in order to answer whether or not this intervention was feasible and impactful at the school level, conducting the research within the school was ultimately the most accessible and reliable method. Collecting data within the school presented its unique challenges, but working with participants in safe, familiar territory compensated for minor inconveniences. Moving forward, refining the methodology included herein by addressing the discussed limitations and recommendations will provide further confidence in the findings as well as the potential to expand the findings to other grades.

Based on the findings herein, teenagers, parents and teachers can begin to see the benefits of incorporating play into their daily regimen. While these findings will benefit from further research and replication, there is a clear relationship between the BBT adolescent curriculum and changes in participant scores pre- and post-intervention. Furthermore, there is a promising trend in the Grade 7 female scores, as they see improvement compared to the males; this warrants further study to better understand the nature of this trend, and whether or not it stems from the hormonally-induced plasticity initiated through puberty. While both of these undertakings are not simple feats, both will greatly add to the current understanding of the nature of the adolescent critical developmental period.

## REFERENCES


Aguilar, D., Blinch, J., & Gonzalez, C. (2021). One brick at a time: Building a developmental profile of spatial abilities. *Developmental Psychobiology*, *63*(6), 1-12. https://doi.org/10.1002/dev.22155

Albuquerque, M. R., Rennó, G. V. C., Bruzi, A. T., Fortes, L. de S., & Malloy-Diniz, L. F. (2022). Association between motor competence and executive functions in children. *Applied Neuropsychology: Child, 11*(3), 495–503. https://doi.org/10.1080/21622965.2021.1897814

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review, 16*(1), 17–42. https://doi.org/10.1007/s11065-006-9002-x

Arnsten, A. F. T. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience, 10*(6), 410–422. https://doi.org/10.1038/nrn2648

Baddeley, A. (1998). The central executive: A concept and some misconceptions. *Journal of the International Neuropsychological Society, 4*(5), 523–526. https://doi.org/10.1017/s135561779800513x

Baghdadi, G., Towhidkhah, F., & Rajabi, M. (2021). Chapter 7 - Assessment methods. In, *Neurocognitive Mechanisms of Attention: Computational Models, Physiology, and Disease States* (pp. 203-250). Academic Press.

Basu, S. (2023). Examining the reliability and validity of computerized Stroop test in children aged 5-13 years: A preliminary study. *Quality & Quantity, 57*(1), 645–653. https://doi.org/10.1007/s11135-022-01376-y

Best, J., & Miller, P. (2010). A developmental perspective on executive function. *Child Development, 81*(6), 1641-1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology, 20*(3), 899–911. https://doi.org/10.1017/S0954579408000436

Blair, C., Granger, D. A., Willoughby, M., Mills-Koonce, R., Cox, M., Greenberg, M. T., Kivlighan, K. T., Fortunato, C. K., & the FLP Investigators. (2011). Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood: Cortisol and cognition. *Child Development, 82*(6), 1970–1984. https://doi.org/10.1111/j.1467-8624.2011.01643.x

Blakemore, S. J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience, 9*(4), 267–277. https://doi.org/10.1038/nrn2353

Blakemore, S. J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry, 47*(3–4), 296–312. https://doi.org/10.1111/j.1469-7610.2006.01611.x

Blakemore, S. J., & Mills, K. L. (2014). Is adolescence a sensitive period for sociocultural processing? *Annual Review of Psychology, 65*(1), 187–207. https://doi.org/10.1146/annurev-psych-010213-115202

Building Brains Together. (n.d.). *Building Brains Together*. Retrieved November 9, 2024, https://www.buildingbrains.ca

Burghardt, G. (2012). Defining and recognizing play. In P. Nathan & A. D. Pellegrini (Eds.), *The Oxford Handbook of the Development of Play.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195393002.013.0002

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports, 100*(2), 126–131. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1424733/

Coelho, L., Amatto, A., Gonzalez, C., & Gibb, R. (2020). Building executive function in pre-school children through play: A curriculum. *International Journal of Play*, *9*(1), 1–15. https://doi.org/10.1080/21594937.2020.1720127

Crews, F., He, J., & Hodge, C. (2007). Adolescent cortical development: A critical period of vulnerability for addiction. *Pharmacology Biochemistry and Behavior, 86*(2), 189–199. https://doi.org/10.1016/j.pbb.2006.12.001

Davis, C. L., Tomporowski, P. D., Boyle, C. A., Waller, J. L., Miller, P. H., Naglieri, J. A., & Gregoski, M. (2007). Effects of aerobic exercise on overweight children's cognitive functioning. *Research Quarterly for Exercise and Sport, 78*(5), 510–519. https://doi.org/10.1080/02701367.2007.10599450.

Davis, C. L., Tomporowski, P. D., McDowell, J. E., Austin, B. P., Miller, P. H., Yanasak, N. E., Allison, J. D., & Naglieri, J. A. (2011). Exercise improves executive function and achievement and alters brain activation in overweight children: A randomized controlled trial. *Health Psychology, 30*(1), 91–98. https://doi.org/10.1037/a0021766

de Bruijn, A. G. M., Hartman, E., Kostons, D., Visscher, C., & Bosker, R. J. (2018). Exploring the relations among physical fitness, executive functioning, and low academic achievement. *Journal of Experimental Child Psychology, 167*(2018), 204–221. https://doi.org/10.1016/j.jecp.2017.10.010

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(2013), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4–12 years old. *Science, 333*(6045), 959–964. https://doi.org/10.1126/science.1204529

Diamond, A., & Ling, D.S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience, 18*(2016), 34–48. https://doi.org/10.1016/j.dcn.2015.11.005

Dishman, R. K., Berthoud, H.-R., Booth, F. W., Cotman, C. W., Edgerton, V. R., Fleshner, M. R., Gandevia, S. C., Gomez-Pinilla, F., Greenwood, B. N., Hillman, C. H., Kramer, A. F., Levin, B. E., Moran, T. H., Russo-Neustadt, A. A., Salamone, J. D., van Hoomissen, J. D., Wade, C. E., York, D. A., & Zigmond, M. J. (2006). Neurobiology of exercise. *Obesity, 14*(3), 345–356. https://doi.org/10.1038/oby.2006.46

Doebel, S., & Zelazo, P. D. (2015). A meta-analysis of the dimensional change card sort: Implications for developmental theories and the measurement of executive function in children. *Developmental Review*, *38*, 241–268. https://doi.org/10.1016/j.dr.2015.09.001

Eberle, S. G. (2014). Toward a philosophy and a definition of play. *American Journal of Play, 6*(2), 214-233. https://eric.ed.gov/?id=EJ1023799

Egan, K. N., Cohen, L. A., & Limbers, C. (2019). Parent–child agreement on the Behavior Rating Inventory of Executive Functioning (BRIEF) in a community sample of adolescents. *Applied Neuropsychology: Child, 8*(3), 264–271. https://doi.org/10.1080/21622965.2018.1438896

Else, P. (2009). *The value of play*. Bloomsbury Publishing. https://www.bloomsbury.com/ca/value-of-play-9780826495655/

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149. https://doi.org/10.3758/BF03203267

Ernst, M., Pine, D. S., & Hardin, M. (2006). Triadic model of the neurobiology of motivated behavior in adolescence. *Psychological Medicine, 36*(3), 299-312. https://doi.org/10.1017/S0033291705005891

Fleischman, C., Perry, R, Hazelwood, V., Pope, F., Gibb, R. & Gonzalez, C.L.R. (2022, September 17). *A play-based curriculum to enhance executive function in adolescents* [Conference poster presentation]. Canadian Undergraduate Research in Psychology Conference, Kelowna, B.C..

Foulkes, L., & Blakemore, S. J. (2018). Studying individual differences in human adolescent brain development. *Nature Neuroscience, 21*(3), 315–323. https://doi.org/10.1038/s41593-018-0078-4

Fuhrmann, D., Knoll, L. J., & Blakemore, S. J. (2015). Adolescence as a sensitive period of brain development. *Trends in Cognitive Sciences, 19*(10), 558–566. https://doi.org/10.1016/j.tics.2015.07.008

Gee, D. G., & Casey, B. J. (2015). The impact of developmental timing for stress and recovery. *Neurobiology of Stress, 1*, 184–194. https://doi.org/10.1016/j.ynstr.2015.02.001

Gershon, R. C., Slotkin, J., Manly, J. J., Blitz, D. L., Beaumont, J. L., Schnipke, D., Wallner-Allen, K., Golinkoff, R. M., Gleason, J. B., Hirsh-Pasek, K., Adams, M. J., & Weintraub, S. (2013). IV. NIH Toolbox Cognition Battery (CB): Measuring language (vocabulary comprehension and reading decoding). *Monographs of the Society for Research in Child Development, 78*(4), 49–69. https://doi.org/10.1111/mono.12034

Gibb, R., Coelho, L., Van Rootselaar, N. A., Halliwell, C., MacKinnon, M., Plomp, I., & Gonzalez, C. L. R. (2021). Promoting executive function skills in preschoolers using a play-based program. *Frontiers in Psychology, 12*(2021)*,* 720225. https://doi.org/10.3389/fpsyg.2021.720225

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., & Rapoport, J. L. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience, 2*(10), 861-863. https://doi.org/10.1038/13158

Ginsburg, K.R., the Committee on Communications & the Committee on Psychosocial Aspects of Child and Family Health. (2007). The importance of play in promoting healthy child development and maintaining strong parent-child bonds. *Pediatrics, 119*(1), 182–191. https://doi.org/10.1542/peds.2006-2697

Gluckman, P. D., Hanson, M. A. (2006). Changing times: The evolution of puberty. *Molecular and Cellular Endocrinology*, *254-255*(2006), 26–31. https://doi.org/10.1016/j.mce.2006.04.005

Gogtay, N., Giedd, J. N, Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., III, Herman, D. H., Clasen, L. S., Toga, A. W., Rapoport, J. L., & Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *PNAS, 101*(21), 8174-8179. https://doi.org/10.1073/pnas.0402680101

Gonzalez, C. L. R., Mills, K. J., Genee, I., Li, F., Piquette, N., Rosen, N., & Gibb, R. (2014). Getting the right grasp on executive function. *Frontiers in Psychology, 5*(285), 1-11. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00285

Gray, P. (2009). Play as a foundation for hunter-gatherer social existence. *American Journal of Play, 1*(4), 476–522. https://eric.ed.gov/?id=EJ1069037

Hillman, C., Erickson, K., & Kramer, A. (2008). Be smart, exercise your heart: Exercise effects on brain and cognition. *Nature Reviews Neuroscience, 9*(1), 58–65. https://doi.org/10.1038/nrn2298

Hubel, D. H., & Wiesel, T. N. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *Journal of Neurophysiology, 26*(6), 994–1002.

Hubel, D. H., & Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *The Journal of physiology, 206*(2), 419–436. https://doi.org/10.1113/jphysiol.1970.sp009022

Ismail, F. Y., Fatemi, A., & Johnston, M. V. (2017). Cerebral plasticity: Windows of opportunity in the developing brain. *European Journal of Paediatric Neurology, 21*(1), 23–48. https://doi.org/10.1016/j.ejpn.2016.07.007

Isquith, P. K., Roth, R. M., & Gioia, G. (2013). Contribution of rating scales to the assessment of executive functions. *Applied Neuropsychology: Child, 2*(2), 125–132. https://doi.org/10.1080/21622965.2013.748389

Jacob, R., & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research, 85*(4), 512–552. https://doi.org/10.3102/0034654314561338

Jaworska, N., & MacQueen, G. (2015). Adolescence as a unique developmental period. J*ournal of Psychiatry & Neuroscience, 40*(5), 291–293. https://doi.org/10.1503/jpn.150268

Jensen, A. R., & Rohwer, W. D. (1966). The Stroop color-word test: A review. *Acta Psychologica*, *25*(1966), 36–93. https://doi.org/10.1016/0001-6918(66)90004-7

Jones, D. R., Dallman, A., Harrop, C., Whitten, A., Pritchett, J., Lecavalier, L., Bodfish, J. W., & Boyd, B. A. (2022). Evaluating the feasibility of the nih toolbox cognition battery for autistic children and adolescents. *Journal of Autism and Developmental Disorders, 52*(2), 689–699. https://doi.org/10.1007/s10803-021-04965-2

Jusko, M. L., Raiker, J. S., Campez, M., Smith, J. N., Fosco, W. D., Horta, L., Little, K., Espinal, K., Sanchez, G., Mattfeld, A. T., Gnagy, E. M., Greiner, A. R., Coles, E. K., & Pelham, W. E. (2021). Brief report: Evaluation of working memory deficits in children with ADHD using the NIH list sorting working memory task. *Child Neuropsychology, 27*(5), 613–620. https://doi.org/10.1080/09297049.2021.1876014

Kassai, R., Futo, J., Demetrovics, Z., & Takacs, Z. K. (2019). A meta-analysis of the experimental evidence on the near- and far-transfer effects among children's executive function skills. *American Psychological Association, 145*(2), 165-88. http://dx.doi.org/10.1037/bul0000180

Kaye, L. K., & Bryce, J. (2014). Go with the flow: The experience and affective outcomes of solo versus social gameplay. *Journal of Gaming & Virtual Worlds, 6*(1), 49–60. https://doi.org/10.1386/jgvw.6.1.49_1

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry, 62*(6), 593. https://doi.org/10.1001/archpsyc.62.6.593

Kolb, B., Whishaw, I. Q., & Teskey, G. C. (2023). *An introduction to brain and behaviour* (7th ed.). Worth Publishers.

Konishi, C., Hymel, S., Zumbo, B. D., & Li, Z. (2010). Do school bullying and student—teacher relationships matter for academic achievement? A multilevel analysis. *Canadian Journal of School Psychology, 25*(1), 19–39. https://doi.org/10.1177/0829573509357550

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *The British Journal of Developmental Psychology, 21*(1), 59-80. https://doi.org/10.1348/026151003321164627

Li, J. W., O'Connor, H., O'Dwyer, N., & Orr, R. (2017). The effect of acute and chronic exercise on cognitive function and academic performance in adolescents: A systematic review. *Journal of Science and Medicine in Sport, 20*(9), 841–848. https://doi.org/10.1016/j.jsams.2016.11.025

Lillard, A. S. (2015). The development of play. In L. S. Liben, U. Müller, & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Cognitive processes* (7th ed., pp. 425–468). John Wiley & Sons, Inc.. https://doi.org/10.1002/9781118963418.childpsy211

Luz, C., Rodrigues, L. P., & Cordovil, R. (2015). The relationship between motor coordination and executive functions in 4th grade children. *European Journal of Developmental Psychology, 12*(2), 129–141. https://doi.org/10.1080/17405629.2014.966073

McLaughlin, K. A., Greif Green, J., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., & Kessler, R. C. (2012). Childhood adversities and first onset of psychiatric disorders in a national sample of US adolescents. *Archives of General Psychiatry, 69*(11), 1151–1160. https://doi.org/10.1001/archgenpsychiatry.2011.2277

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

National Institutes of Health. (2020). *NIH Toolbox® for assessment of neurological and behavioral function administrator's manual*. National Institutes of Health and Northwestern University. https://nihtoolbox.zendesk.com/hc/en-us/articles/13364439509524-NIH-Toolbox-Admin-Manuals

NIH Toolbox. (n.d.-a). *Dimensional Change Card Sort Test.* Retrieved November 13, 2024, from https://nihtoolbox.org/test/dimensional-change-card-sort-test/

NIH Toolbox. (n.d.-c). *Picture Vocabulary Test*. Retrieved November 13, 2024, from https://nihtoolbox.org/test/picture-vocabulary-test//

Pellis, S. M., & Pellis, V. C. (2007). Rough-and-tumble play and the development of the social brain. *Current Directions in Psychological Science, 16*(2), 95-98. https://doi.org/10.1111/j.1467-8721.2007.00483.x

Pellis, S. M., & Pellis, V. C. (2009). *The playful brain: Venturing to the limits of neuroscience*. Oxford, England: Oneworld.

Pellis, S. M., Pellis, V. C., & Bell, H. C. (2010). The function of play in the development of the social brain. *American Journal of Play, 2*(3), 278–298. https://eric.ed.gov/?id=EJ1069225

Perrin, J. S., Leonard, G., Perron, M., Pike, G. B., Pitiot, A., Richer, L., Veillette, S., Pausova, Z., & Paus, T. (2009). Sex differences in the growth of white matter during adolescence. *NeuroImage, 45*(4), 1055–1066. https://doi.org/10.1016/j.neuroimage.2009.01.023

Piaget, J. (1951). *Play, Dreams And Imitation In Childhood*. Routledge. https://doi.org/10.4324/9781315009698

Poon, K. (2018). Hot and cool executive functions in adolescence: Development and contributions to important developmental outcomes. *Frontiers in Psychology, 8*(2311). https://doi.org/10.3389/fpsyg.2017.02311

Rigoli, D., Piek, J. P., Kane, R., & Oosterlaan, J. (2012). An examination of the relationship between motor coordination and executive functions in adolescents. *Developmental Medicine & Child Neurology, 54*(11), 1025–1031. https://doi.org/10.1111/j.1469-8749.2012.04403.x

Ruipérez-Valiente, J. A., & Kim, Y. J. (2020). Effects of solo vs. collaborative play in a digital learning game on geometry: Results from a K12 experiment. *Computers & Education, 159*(104008). https://doi.org/10.1016/j.compedu.2020.104008

Sala, G., & Gobet, F. (2016). Do the benefits of chess instruction transfer to academic and cognitive skills? A meta-analysis. *Educational Research Review, 18,* 46–57. https://doi.org/10.1016/j.edurev.2016.02.002

Sawyer, S. M., Azzopardi, P. S., Wickremarathne, D., & Patton, G. C. (2018). The age of adolescence. *Lancet Child Adolescent Health, 2*(3). 1-6. http://dx.doi.org/10.1016/S2352-4642(18)30022-1

Selemon, L. D. (2013). A role for synaptic plasticity in the adolescent development of executive function. *Translational Psychiatry, 3*(3). https://doi.org/10.1038/tp.2013.7

Sember, V., Jurak, G., Kovač, M., Morrison, S. A., & Starc, G. (2020). Children's physical activity, academic performance, and cognitive functioning: A systematic review and meta-analysis. *Frontiers in Public Health, 8*(307). https://doi.org/10.3389/fpubh.2020.00307

Shaheen, S. (2014). How child's play impacts executive function–related behaviors. *Applied Neuropsychology: Child, 3*(3), 182–187. https://doi.org/10.1080/21622965.2013.839612

Shephard, R. J. (1997). Curricular physical activity and academic performance. *Pediatric Exercise Science, 9*(2), 113-126. https://doi.org/10.1123/pes.9.2.113

Smart Moves. (n.d.). *Crackabout*. Retrieved on July 2, 2024 from https://www.smartmoves.ai/step/view/6603392f2cd76ac0002eeb5a

Sochat, V. (n.d.) *Stroop-5min*. The Experiment Factory. Retrieved November 12, 2024 from https://expfactory-experiments.github.io/stroop-5min/

Sochat, V. (2018). The Experiment Factory: Reproducible experiment containers. *Journal of Open Source Software*, *3*(22), 521, https://doi.org/10.21105/joss.00521

Sochat, V. V., Eisenberg, I. W., Enkavi, A. Z., Li, J., Bissett, P. G., & Poldrack, R. A. (2016). The Experiment Factory: Standardizing behavioral experiments. *Frontiers in Psychology*, *7*(610). https://doi.org/10.3389/fpsyg.2016.00610

Sointu, E. T., Savolainen, H., Lappalainen, K., & Lambert, M. C. (2017). Longitudinal associations of student–teacher relationships and behavioural and emotional strengths on academic achievement. *Educational Psychology, 37*(4), 457–467. https://doi.org/10.1080/01443410.2016.1165796

Somerville, L. H., Bookheimer, S. Y., Buckner, R. L., Burgess, G. C., Curtiss, S. W., Dapretto, M., Elam, J. S., Gaffrey, M. S., Harms, M. P., Hodge, C., Kandala, S., Kastman, E. K., Nichols, T. E., Schlaggar, B. L., Smith, S. M., Thomas, K. M., Yacoub, E., Van Essen, D. C., & Barch, D. M. (2018). The lifespan human connectome project in development: A large-scale study of brain connectivity development in 5–21 year olds. *NeuroImage*, 183, 456–468. https://doi.org/10.1016/j.neuroimage.2018.08.050

Somerville, L. H., Jones, R. M., & Casey, B. J. (2010). A time of change: Behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. *Brain and Cognition, 72*(1), 124–133. https://doi.org/10.1016/j.bandc.2009.07.003

Steinberg, L. (2010). A dual systems model of adolescent risk-taking. *Developmental Psychobiology, 52*(3), 216–224. https://doi.org/10.1002/dev.20445

Stuss, D. T., Floden, D., Alexander, M. P., Levine, B., & Katz, D. (2001). Stroop performance in focal lesion patients: Dissociation of processes and frontal lobe lesion location. *Neuropsychologia, 39*(8), 771–786. https://doi.org/10.1016/S0028-3932(01)00013-6

Tamnes, C. K., Herting, M. M., Goddings, A.L., Meuwese, R., Blakemore, S.J., Dahl, R. E., Güroğlu, B., Raznahan, A., Sowell, E. R., Crone, E. A., & Mills, K. L. (2017). Development of the cerebral cortex across adolescence: A multisample study of inter-related longitudinal changes in cortical volume, surface area, and thickness. *The Journal of Neuroscience, 37*(12), 3402–3412. https://doi.org/10.1523/JNEUROSCI.3302-16.2017

Tooley, U. A., Bassett, D. S., & Mackey, A. P. (2021). Environmental influences on the pace of brain development. *Nature Reviews Neuroscience, 22*(6). https://doi.org/10.1038/s41583-021-00457-5

Tulsky, D. S., Carlozzi, N. E., Chevalier, N., Espy, K. A., Beaumont, J. L., & Mungas, D. (2013). V. NIH Toolbox Cognition Battery (CB): Measuring working memory. *Monographs of the Society for Research in Child Development, 78*(4), 70–87. https://doi.org/10.1111/mono.12035

Tulsky, D. S., Carlozzi, N., Chiaravalloti, N. D., Beaumont, J. L., Kisala, P. A., Mungas, D., Conway, K., & Gershon, R. (2014). NIH Toolbox cognition battery (NIHTB-CB): List sorting test to measure working memory. *Journal of the International Neuropsychological Society, 20*(6), 599–610. https://doi.org/10.1017/S135561771400040X

Walker, J., & D'Amato, R. (2006). Test review: Behavior Rating Inventory of Executive Function–Self-Report version. *Journal of Psychoeducational Assessment*, *24*(4). https://doi.org/10.1177/0734282906288390

Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press.

Yogman, M., Garner, A., Hutchinson, J., Hirsh-Pasek, K., Golinkoff, R. M., Committee on Psychosocial Aspects of Child And Family Health, Council On Communications And Media, Baum, R., Gambon, T., Lavin, A., Mattson, G., Wissow, L., Hill, D. L., Ameenuddin, N., Chassiakos, Y. (Linda) R., Cross, C., Boyd, R., Mendelson, R., Moreno, M. A., Radesky, J.,  Swanson, W.S., Hutchinson, J., Smith, J. (2018). The power of play:

A pediatric role in enhancing development in young children. *Pediatrics, 142*(3), 1-16. https://doi.org/10.1542/peds.2018-2058

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives, 6*(4), 354–360. https://doi.org/10.1111/j.1750-8606.2012.00246.x

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development, 78*(4), 16–33. https://doi.org/10.1111/mono.12032

Zysset, A. E., Kakebeeke, T. H., Messerli-Bürgy, N., Meyer, A. H., Stülb, K., Leeger-Aschmann, C. S., Schmutz, E. A., Arhab, A., Puder, J. J., Kriemler, S., Munsch, S., & Jenni, O. G. (2018). Predictors of executive functions in preschoolers: Findings from the SPLASHY study. *Frontiers in Psychology, 9*(2018)*,* 1-11. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02060

## APPENDIX A: ETHICS APPROVAL

UNIVERSITY OF ALBERTA

ARISE
Alberta Research Information Services

Date: October 27, 2022 2:35:00 PM

Print Close

# Table of Contents