The Effect of Adaptive Difficulty Adjustment on the Effectiveness of a Game to Develop Executive Function Skills for Learners of Different Ages

Jan L. Plass*, Bruce D. Homer**, Shashank Pawar*

Corinne J. Brenner*, & Andrew P. MacNamara*

* CREATE Lab, New York University

** The Graduate Center, City University of New York

Published in 2019 in *Cognitive Development*, 49, 56–67.

Link to Peer-Review Process

Abstract

Research suggests that gains in executive function (EF) skills training are strongest when task difficulty increases progressively, yet findings on the effectiveness of adaptive approaches for EF training are inconsistent. This study compared the effectiveness of an adaptive vs a non-adaptive version of a digital game designed to train the EF sub-skill of shifting. Results showed increases in shifting skills for all learners between pretest and posttest measures, with adolescents scoring higher than pre-adolescents and early adolescents on posttest measures. Data analysis uncovered a trend suggesting that the adaptive treatment may be more effective than the non-adaptive treatment for adolescents. User logs showed that adaptivity helped customize players' gameplay based on their performance, by making game play easier for younger learners, and making game play more difficult for older learners. Results support the use of digital games to train EF for a broad range of learners.

Keywords: Executive functions; Adaptivity; Game-based learning; Zone of Optimal Engagement

The Effect of Adaptive Difficulty Adjustment on the Effectiveness of a Digital Game to Develop Executive Function Skills for Learners of Different Ages

Can digital games help children develop executive function skills, and what importance does adaptivity have for the game's effectiveness? Digital games are interactive software programs with incentive systems, clear goal states, and immediate feedback. They can be designed to involve tasks that place demands on targeted cognitive resources that are tapped by specific cognitive processes, such as executive functions. For example, in a game targeting cognitive flexibility, which is our ability to switch between different tasks or mental states, the player would receive specific rules on how to play. On a screen where game characters called *Aliens* move from their space ship toward Earth because they are hungry, the rules would describe which type of *Alien* game character preferred food, and which preferred drinks. These rules would change in subsequent trials, and the changes would be increasing both in frequency and complexity as the game progressed.

Games have been considered a potentially useful and effective medium for learning and cognitive development because game features can be designed to motivate players, facilitate a broad range of tasks at different levels of complexity, and record the player's action in a log file for future analysis. Researchers have investigated games that aimed to enhance socio-emotional variables (Hromek & Roffey, 2009), study skills (Charlton, Williams & McLaughlin, 2005; Ke, 2008), and cognitive skills (Powers et al., 2013) for learners of different ages. Among cognitive skills, researchers and practitioners are especially interested in the development of executive function (Diamond & Lee, 2011). Executive functions (EF) are cognitive processes that are used for effortful, controlled, and goal-directed thinking and behavior (Banich, 2009; Best, 2012).

Learners use EF skills to manage their cognitive processing when they aim to achieve a goal. EF skills are significant because they have been shown to predict academic success (Best, 2014; St Clair-Thompson & Gathercole, 2006), metacognitive skills (Bryce, Whitebread & Szűcs, 2014), language acquisition skills (Fuhs et al., 2014), and theory of mind (Carlson, Moses & Breton, 2002), among other skills.

The goal of this study was to determine whether the effectiveness of a game specifically designed to train EF skills can be increased by adding a feature that adaptively adjusts the difficulty level of the EF training tasks to the current ability levels of the individual. For this purpose, we utilize a previously validated game for EF training. The *Alien Game* was extensively tested in the lab and in field settings in schools with players aged 10-17 years who had a broad range of cognitive abilities (Plass, 2016). Cognitive consequences research has shown that this game is able to enhance players' shifting skill, as measured by the Dimensional Change Card Sorting task (Zelazo, 2006) after approximately 1.5 hours of play for younger children (Homer, Plass, Rafaele, Ober, & Ali, 2018) and after approximately 2 hours of play for college students (Parong, Mayer, Fiorella, MacNamara, Homer, Plass, 2017). In this experiment, we compared an adaptive and a non-adaptive version of the Alien game. We were especially interested in examining if adaptivity results in different play experiences for learners of different ages, and how these experiences, would affect EF gains. Insights about the influence of adaptivity of game-based training would validate adaptivity as a key design feature for EF games and would allow EF intervention designers to enhance the effectiveness of their programs for learners of different ages.

Theoretical Framework

Executive Function

Executive functions (EF) are a set of cognitive skills required to plan, monitor and control cognitive processes. EF, which develop throughout childhood into adolescence and early adulthood (Müller & Kerns, 2015), are related to a number of important outcomes, including behavioral problems (Barkley, 1997), social functioning (Riggs et al., 2006), and academic success (Fuhs et al., 2014; Müller & Kerns, 2015). There is also considerable evidence for the importance of EF for school readiness. For example, Blair & Razza (2007) found that EF skills in preschool uniquely predicted children’s math and literacy skills in kindergarten.

The literature documents different theoretical models for EF, however, the unity/ diversity model (Miyake, Friedman, Emerson, Witzki, Howerter & Wager, 2000; Miyake & Friedman, 2012) is especially fitting for the purpose of this study as it provides support for a non-unitary distributed structure of EF. This structure allows development of targeted game tasks for enhancing specific EF subskills. According to this model, EF is comprised of three closely related but distinguishable component skills: updating, shifting, and inhibition. *Updating* is defined as updating and monitoring of working memory representations. It entails the accommodation of new task-relevant information by replacing old irrelevant information held in working memory. A common task requiring updating skills is the keep-track task (Yntema, 1963). In this task, participants are shown a sequence of stimuli of different categories like fruits, vegetables and animals. The goal for the participants is to remember the last stimuli of each category, i.e. the last seen fruit, vegetable and animal. The need to update items in working memory in real-time taps into the updating skills of participants. *Shifting* is defined as shifting between tasks or mental sets. It is the ability to switch from one task or perspective to another.

For example, in the local-global task (Navon, 1977), participants are presented with large letter made from smaller letters (example: an ‘H’ made from small l’s). During the task, the participants must switch perspective from local (answering ‘l’) to global (answering ‘H’). The perspective switching required in this task requires shifting skills. *Inhibition* is defined as the suppression of a dominant or prepotent response and is the ability to control our attention, behaviors, thoughts, and emotions. It requires deliberate suppression of intuitive reactions while performing a task. For example, in the Stroop task (Stroop, 1935), participants are requested to choose one of the two color blocks based on either the text or the font color of the presented color-word. For example, when presented with the word ‘red’ but in blue font color, the correct answer when asked to follow text is red, but the correct response when following font color is blue. When performing this task, there is a difference in reaction time when following font color compared to following the text (Stroop, 1935). This difference occurs because, following font color requires suppression of automatic reading of the text. The intuitive response is to refer to the text ‘red’ rather than the font color blue. Overcoming this intuitive response requires inhibition skills. In the framework of the unity-in-diversity model, this study is focused on the development of shifting skills.

Development of Executive Functions

Because of the importance of EF described above, there has been a considerable amount of research on the development of EF, and scholars as well as practitioners have shown great interest in creating successful interventions for improving EF. EF skills develop over time, from early childhood to adulthood, but this development is not linear and happens in spurts, with different components of EF developing at different rates (Carlson, Zelazo, & Faja, 2013; Zelazo, 2006). As younger children have lower EF, training tasks for them need to be designed to create less of a challenge compared to EF training tasks for individuals who are closer to adulthood and therefore closer to the full development of EF. As we get older, gains are slower and greater challenges are required to see training benefits.

Among the interventions that have been used to help learners develop EF skills, digital games have been highlighted as a particularly promising medium (Green & Bavelier, 2008). Video games have many of the features that researchers identified as important for EF skills improvement, such as the ability of engaging players for long periods of time (Kahne et al., 2008), providing timely and valuable feedback (Shute, Rieber & Van Eck, 2011), offering attractive rewards (Green & Bavelier, 2008), and embedding emotional design using characters and narrative elements (Szczuka et al., 2013; Plass, 2017). These features are desirable for training tasks (Green & Bavelier, 2008) and make video games a prime candidate for EF training.

These inherent affordances of video games, combined with the rising popularity of games (Best The News, 2016; Entertainment Software Association, 2016), have encouraged the development of a number of video game interventions for EF skills training (Nouchi et al., 2012; McNab et al., 2015; Alloway & Alloway, 2008; Van der Molen & Luit, 2010). Even though several studies found digital games to be successful in helping learners develop EF skills (Parong et al., 2017; Homer et al., 2017; Klingberg et al., 2005; Persson & Reuter-Lorenz, 2008), a meta-analysis of the effect of EF interventions on a range of cognitive skills has reported mixed results regarding the efficacy of games for training EF (Powers et al., 2013). These inconsistencies may be associated with design features of the games used for the interventions (Enriquez-Geppert, Huster & Herrmann, 2013), and for the fact that the analysis focused on EF as a unitary construct rather than on the sub-skills of EF described above.

One possible explanation for these mixed results is in the approach that has been taken to designing games as EF interventions. One of these approaches has been described as “brain training” games, digital applications that are often merely gamified versions of EF tasks. In other words, these gamified versions of EF tasks use existing EF measures and merely add game features such as rewards and feedback systems to enhance motivation. Since they do not alter the training task, these gamified versions utilize an assessment task for the purpose of training EF. Studies investigating the effectiveness of gamified tasks have found mixed results. One study found significant EF improvements in older adults who played the brain-training game, *Brain Age*. During this study, participants played 8 different training games focused on arithmetic calculation, reading tasks, sorting tasks, counting tasks, and geometric tasks. The study was conducted for 15 minutes a day, 5 days per week, over a 4-week period, as compared to a control group that played the game Tetris (Pajitnov, 1989), following the same schedule (Nouchi et al., 2012). However, other studies have not found significant effects. One study investigated the effectiveness of one of the commercially most successful EF games, *Lumosity*. This game also consists of several mini-games categorized as speed games, memory games, attention games, and flexibility games, among others. This study found that even after 15–20 hours of play, few, if any, cognitive benefits existed (Bainbridge & Mayer, 2017). Another study found that participants who were trained with digital EF tasks got better only at those specific tasks–the effects did not transfer to other tasks (Van Muijden, Band, & Hommel, 2012).

In cases where different research studies come to different conclusions about the ability of an intervention to have a specific effect, three general types of explanations are conceivable: One is that the target skill cannot be in fact trained; a second, that specific features need to be present for the intervention to be successful, and a third, that the effectiveness of the intervention depends on specific characteristics of the learner. The lack of support for the effectiveness of gamified tasks may be due to the fact that tasks that are useful for the measurement of EF may not necessarily be effective in developing EF. Given the large number of studies indicating that EF skills can in fact be trained (Diamond & Lee, 2011; Karbach & Verhaeghen, 2014), we pursued in the studies presented here the latter two possibilities, focusing on the question whether adaptive EF games are better able to support learners of different ages than non-adaptive games. The games used for this research were specifically designed for the purpose of supporting EF development, rather than gamifications of EF measures, see details in the method section below.

Adaptivity and EF Training

Adaptive systems are designed to cater to individual differences of users. Adaptivity is here defined as the ability of an intervention to provide each individual with the kind of experience she or he needs at any given time in order to be successful in reaching the intended outcomes (Plass, 2016). Studies have shown that factors such as prior knowledge (Alexander & Judy, 1988; Tobias, 1994), emotional states such as frustration, boredom, motivation, and confidence (Craig, Graesser, Sullins, & Gholson, 2004; Qu, Wang, & Johnson, 2005), and differences in demographic and sociocultural factors (Conchas, 2006; Desimone, 1999) are among the predictors of learning outcomes. To address these and other learner differences, adaptive systems can optimize game parameters at a cognitive, affective, socio-cultural or motivational level (Plass, 2016).

Our present research was concerned with adaptivity that targeted cognitive factors, specifically, the adaptive adjustment of the difficulty of the training task. In order to support the development of EF skills for learners of different ages, the game needs to provide tasks with a difficulty level matching the current EF abilities of the user, and needs to be engaging enough so players continue completing these tasks. The idea that adaptivity will be able to enhance the efficacy of EF interventions is based on Diamond and Lee's (2011) review of effective EF interventions, in which they suggest that effects are strongest for interventions requiring substantial executive control, and for which difficulty is progressively increasing (Holmes, Gathercole & Dunning, 2009; Klingberg et al., 2005). Related research suggests that EF training yields better outcomes when the task difficulty is easy at first and is increased with time (Ahissar & Hochstein, 2004). When learners start performing a task, their skills gradually improve and tasks which were initially difficult eventually become trivial and disengaging. As a result, the current skill level of learners determines whether a training task of certain difficulty will have cognitive benefits or not. Consequently, if a task is too easy, the learner's cognitive resources are not taxed sufficiently. If the task is too difficult, the learners' resources are taxed too much. In both cases, the learner may become disengaged from the task (Lövdén, Bäckman, Lindenberger, Schaefer & Schmiedek, 2010), which is undesirable from a motivational and an EF training perspective. In order to keep the learner engaged in the kind of task that is most likely to develop their EF, the game must adjust the task difficulty based on the learner's skill level and accommodate changes in these skill levels. We describe this process as allowing the learner to maintain an *optimal zone of engagement* for cognitive skills training.

Zone of Optimal Engagement

We use the concept of the *Zone of Optimal Engagement* to describe the goal of an EF intervention to create a state of high cognitive engagement. In this state, the learner exerts a high level of executive control, and difficulty levels are maintained to assure that the challenges and tasks are always in the right range of difficulty for this individual. This means the task is difficult enough so the player does not disengage due to feeling bored, and not too difficult, so the player does not disengage due to feeling overwhelmed. This concept of a Z*one of Optimal Engagement* follows the features for effective EF interventions identified by Diamond and Lee (2011), Holmes, Gathercole, and Dunning (2009), and Klingberg et al. (2005). These investigations have shown that effects are strongest when interventions require substantial executive control, and when difficulty progressively increases. We therefore consider the idea of keeping the learner within their Zone of Optimal Engagement a key ingredient for effective interventions to develop EF. The Zone of Optimal Engagement goes beyond recognizing optimal level of challenge for each player and identifies a range of difficulty beneficial for training. Further, it acknowledges this range to be a moving target, as learners' skills improve with training. Since this usually means that a high amount of cognitive resources need to be expended, the game is hard for players, and motivational game features are used to compel the player to continue engaging with the task by making the game fun to play.

The adjustment of the difficulty level to keep the individual in the *Zone of Optimal Engagement* can be achieved using an adaptive algorithm[1] that, based on an individual's performance, adjusts the task demands specifically for this individual. The research described above has shown such an adaptive difficulty adjustment plays a pivotal role in cognitive training tasks. An intervention that adapts to the learners' need should be able to adjust the task difficulty based on the needs of the learner. However, despite the theoretical support for such adaptive difficulty adjustment as a design feature for interventions (Diamond & Lee, 2011; Holmes, Gathercole, & Dunning, 2009; Klingberg et al., 2005), existing empirical evidence is inconclusive as to its effectiveness (Von Bastian, & Eschen, 2016). Reviews suggest that this

[1] An algorithm is here a set of rules to be followed that compute the optimal level of difficulty for a learner based on the learner's performance.

may in part be due to the lack of clarity in defining the specific function and purpose of adaptive systems used (Holmes et al., 2009; Morrison, & Chein, 2011; Von Bastian, & Eschen, 2016). Often, the definition of what is meant by *adaptive* varies greatly among researchers. This study aims to contribute to our understanding of the effect of adaptive difficulty adjustment in the game progression on improving EF training outcomes in a game-based training context and for learners of different age ranges.

Methodology

Treatment design

This study investigates the effect of adaptive difficulty adjustment on the enhancement of EF skills through a game-based intervention. We implement a game-based intervention as reviews of effective EF interventions have shown that participants' willingness to devote time and task motivation are important factors for an intervention to be successful (Diamond & Lee, 2011). The game used in the present study is the *Alien Game*, which was specifically developed for research purposes (CREATE, 2015). In this game, players must feed aliens the correct item (food or drink) before they descend to the bottom of the screen (see Figure 1, left panel). Aliens vary in terms of color (red or blue) and number of eyes (one or two). The feeding rules of what each type of alien needs to be given, e.g. food for two-eyed aliens and drinks for one-eyed aliens, are shown to players before each round begins (see Figure 1, right panel). Each level consists of multiple rounds that each have different feeding rules that increase in complexity.

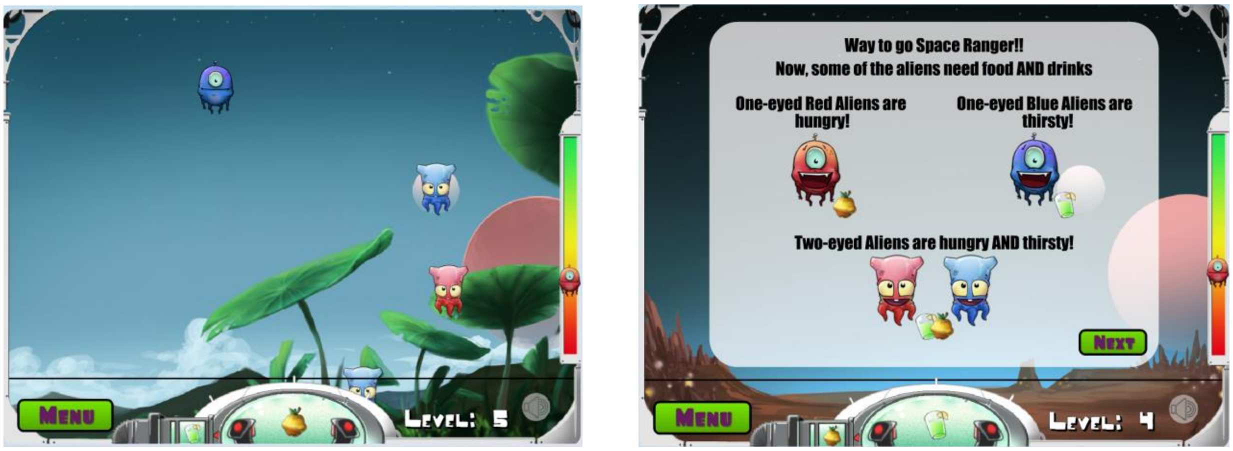

Figure 1. The Alien Game: Gameplay and Feeding Rules (CREATE, 2015)

Switching between feeding requirements as the various types of aliens appear, coupled with the need to switch to new feeding rules in each round, was specifically designed to train the EF skill of shifting. More advanced levels introduce feeding rules - mechanics[2] that make individual aliens and/or all aliens temporarily want the opposite of their current feeding rule, further increasing the need for the player to utilize their EF sub-skill of shifting to succeed in the game. A typical gameplay sequence for the *AlienGame* would unfold as follows:

1. The player learns the feeding rule: Red aliens are hungry; Blue aliens are thirsty.
2. The player internalizes this rule and starts feeding red aliens with food and blue aliens with drinks.
3. After a few minutes of gameplay, a rule switch happens. The new feeding rule may be as follows: one-eyed aliens are hungry and two-eyed aliens are thirsty.
4. This switch requires the use of shifting skills as the player must resist the urge to apply the old rule and shift their mental set to focus on the new rule, involving new characteristics (number of eyes instead of color of the aliens) and play accordingly.

[2] Game mechanics are the essential building blocks of a game, the core tasks that players repeat throughout the game play (Salen & Zimmerman, 2004).

The purpose of adaptivity in the present research is to adjust the difficulty of the training task to facilitate optimal cognitive engagement and to avoid disengagement and frustration. This is achieved by manipulating game difficulty in real-time based on players' gameplay performance. The adaptive engine in our EF training game takes real time in-game performance as input and modifies the training task to maintain task difficulty at a level that provides optimal cognitive engagement. To create an adaptive engine for the *Alien Game*, we reviewed the potential design factors that could be used for difficulty adjustment. There are several game variables in this specific game that can be modified to adjust the difficulty of the game:

- *Number of Aliens*: the number of Aliens visible on screen at any given time; a result of the speed of spawning new Aliens,
- *Speed of Aliens*: the numbers of pixels per second with which Aliens move from the top to the bottom of the screen,
- *Frequency of rule changes*: the frequency with which new rules are presented that specify which particular types of Aliens like food versus drink.
- *Complexity of Rules*: the complexity of the rules guiding Aliens' food or drink preferences, e.g., based on their color, the number of their eyes, or whether they were hit by lighting or radiation.

Because of these multiple options of adjusting the difficulty of the game, a nested approach is required that determines which of these variables is changed at what stage in the game to keep the learner in the Zone of Optimal Engagement. Therefore, some variables are only adjusted when a new level is reached, such as the introduction of more complex rules, whereas others can be adjusted within a given level, such as the speed of the Aliens. Since no research exists that could guide decision of how to structure difficult adjustments in a game, these decisions are often made by game designers. Table 1 shows the approach used for the present study.

Table 1. Nested Approach to Adaptivity in Alien Game

Variable	**Stage of Adjustment in the *Alien Game***
Complexity of feeding rules	Between game levels
Frequency of feeding rule changes	Between game levels
Number of Aliens present on screen	Within game level–higher priority
Speed of Aliens moving down the screen	Within game level–lower priority; between levels

As this table shows, the two variables that are used within each level to dynamically adjust the difficulty of the EF tasks are *Number of Aliens* and *Speed of Aliens*. Following other cognitive skills training tasks, we used the following rules for the difficulty adjustments:

A. If the player completes three consecutive tasks, i.e. feeds three consecutive aliens the correct item successfully, increase game difficulty by increasing number of aliens on screen and increasing the fall speed of aliens.

B. If the player fails to complete a task successfully, i.e. feeds an alien the incorrect item or misses to feed an alien before it disappears from the screen, decrease difficulty by decreasing the number of aliens on screen and decreasing the fall speed of aliens.

The research design addressed the following research questions:

1. Is the game effective for the development of learners' executive functions as measured by changes in the DCCS task (Zelazo, 2006)?
2. Does the adaptive adjustment of difficulty levels in the EF game result in larger improvements of shifting skills than difficulty increases that are fixed for all learners?

3. Do play outcomes and the effect of adaptivity depend on participants' age or other demographic variable?

We conducted an experiment with middle school and high school students to answer these questions.

Participants and Research Design

Middle and high school students (*N* = 101) from a large urban city in the Northeastern United States participated in this lab-based study. 16 participants were excluded from the analyses due to missing data, leading to a dataset of 85 participants. An additional 10 participants had to be excluded as they had too many incorrect responses on the DCCS, for a final sample size of 75 students. The participants were between the ages of 10 to 17 years (*M*=12.13, *SD*=1.71), and 35 participants identified as female. The study was approved by the IRB panels of the city's department of education and of the university. All participants provided parental/guardian permissions before arriving at the lab and were also consented (or, if younger than 13 years, assented) in person during the visit. This study used a randomized control design with an active control group playing the non-adaptive version of *Alien Game* and the treatment group playing the adaptive version.

Materials and Apparatus

The *Alien Game,* hosted on the online experiment manager DREAM (CREATE, 1996), was used as intervention in this study. Two versions of this game, one adaptive and one non-adaptive, were used as treatment and control conditions, respectively. The two versions were identical except for the method used for difficulty adjustment. In the adaptive version, the game difficulty increased or decreased based on player performance. As described above, an increase or decrease in difficulty was triggered after three consecutive correct responses, or after each incorrect response, respectively. In the non-adaptive version, difficulty increases in the speed and complexity of each level were designed based on a series of user research studies and were fixed for all participants–no adaptive changes occurred.

Demographics. A short demographic questionnaire was used to collect information about participants' age, gender, and educational background including their grade level and their guardians' highest level of education. The survey also included questions about participants' gameplay experience. For example, participants were asked how many hours they played video games on a typical school day. This information was used to verify that any differences found were not attributable to factors such as game play experience. The survey was administered digitally using the Qualtrics platform and was hosted on the DREAM system.

EF Skills. We used a digital version of the Dimensional change card sorting task (DCCS) to measure EF skills of participants as pre- and post-test. Our version of DCCS was based on the NIH guidelines as discussed by Zelazo and Bauer (2013) The DCCS is a validated and widely accepted task for assessing EF skills (Zelazo, 2006). In this task, participants sort image cards with visuals that differ along two dimensions, color and shape. The participants are first asked to sort cards according to one of the dimensions (color). Next, participants are asked to sort the same set of cards, but according to the other dimension (shape). Finally, after these practice tasks, participants receive stimuli along with prompts regarding which dimension should be used for sorting (color or shape). The DCCS version used in this study included four blocks: a practice block with 8 trials (4 shape trials and 4 color trials) that provided feedback about correct and incorrect choices; a pre-switch block that included 5 trials on either the shape or color dimension and provided no feedback, but would be repeated until participants responded correctly to at least

4 trials; a post-switch block of 5 trials that would explore the dimension that was not explored in the pre-switch block, and would also repeat if fewer than 4 responses were correct; and a mixed block with 30 trials, with 23 shape trials and 7 color trials, where no feedback was provided and dimensions were switched between trials. The scoring was also conducted using NIH guidelines (Zelazo & Bauer, 2013). The score range for this task was between 0 and 10, and floor or ceiling effects were not observed with our participants (top 25% of scores ranged between 7.75 and 8.74).

Game User Logs. The game recorded all player actions as well as game events, which were time stamped with millisecond precision. Events were recorded each time an alien was given correct food or drink, incorrect food or drink, or not fed at all (missed response). Also recorded were all difficulty adjustments by the adaptive engine. This information was utilized to generate gameplay statistics describing the effect of adaptive changes on gameplay. For example, statistics such as the count of difficulty increasing and decreasing changes made by the adaptive engine were generated. These logs were stored by the DREAM system (CREATE, 1996) in a database on a secure server.

Procedure

During the visit, participants were first randomly assigned to one of the experimental conditions. Next, they completed the consent materials as approved by the review boards of [blinded university], [blinded university], and [blinded City Department of Education]. All participants finished the demographics survey followed by the DCCS pretest. Then, the participants played either the adaptive or the non-adaptive version of the *Alien Game* for 20 minutes. Finally, participants completed the DCCS posttest, and a gameplay questionnaire including questions evaluating their gameplay experience.

Results

To address **research question 1**, which asked whether the game was effective overall, we conducted a one-tailed paired t-test to investigate changes in DCCS scores before and after the intervention. The mean pretest and posttest DCCS scores were 6.60 (*SD*= 1.52) and 6.97 (*SD* = 1.18), respectively. Results from the analysis showed paired sample t test: $t(74) = 1.86$, $p = .033$ (one-tailed), $d = 0.22$.

To address **research question 2**, which asked whether the adaptive adjustment of difficulty levels in an EF game results in larger improvements of shifting skills than difficulty increases that are fixed for all learners, we conducted a one-way ANCOVA with posttest DCCS scores as dependent variable and pretest DCCS scores as a covariate. We did not find a significant main effect of treatment on DCCS posttest scores ($F(1, 72) = .77$, $p = .38$, $M_{Post} = 7.09$, $SE_{Post} = .199$ for the adaptive, $M_{Post} = 6.86$, $SE_{Post} = .181$ for the non-adaptive game).

To address **research question 3**, which asked whether the treatment effects differed for different age groups, we conducted a repeated measures analysis of variance (ANOVA) with two factors to analyze the effects of age and treatment on pretest versus posttest DCCS scores, see table 2 for descriptives. The first factor was condition (adaptive, non-adaptive). The second factor was age group. Three age groups were created based on developmental stage definitions of late childhood/pre-adolescence, early adolescence, and adolescence: one for players 10-11 years old, one for players 12-13 years old, and one for players 15 years and older. Here, we observed a significant main effect of time, such that posttest scores were higher on average than pretest scores $F(1, 69) = 4.07$, $p = .048$, $MSE = 6.17$, $\eta_p^2 = 0.056$. We also observed a significant main effect of age group, $F(2, 69) = 5.48$, $p = .006$, $MSE = 10.90$, $\eta_p^2 = 0.14$. The interaction effect of age and treatment on DCCS scores was not statistically significant, $F(2, 69) = .87$, $p = .42$, $MSE = 1.725$, $\eta_p^2 = 0.025$. Means and standard deviation of DCCS pre- and posttest scores by condition and age group are displayed in table 2.

For the main effect of age group, post-hoc LSD comparisons of differences between DCCS scores revealed that players 15 years and older had statistically significantly larger mean differences than players 10-11 years old (*Mean difference* = .90, *SE* = .285, *p* = 0.002), and significantly higher than players 12-13 years old (*Mean difference* = .91, *SE* = .35, *p* = 0.011). No other statistically significant differences were observed.

Table 2: Mean (SD) of DCCS pretest and posttest scores by age group

Age Group	**N**	**Pre-test Mean (SD)**	**Post-test Mean (SD)**
10-11 years	41	6.43 (1.38)	6.71 (1.16)
Non-adaptive	25	6.33 (1.38)	6.58 (1.25)
Adaptive	16	6.59 (1.40)	6.91 (1.00)
12-13 years	16	6.17 (1.75)	6.95 (1.06)
Non-adaptive	6	6.10 (2.02)	7.29 (.78)
Adaptive	10	5.97 (1.79)	6.74 (1.19)
15+ years	18	7.36 (1.42)	7.56 (1.15)
Non-adaptive	10	7.09 (1.24)	7.20 (1.29)
Adaptive	8	7.70 (1.63)	8.00 (.84)

For the interaction of age group and treatment, post-hoc comparisons of differences between DCCS scores for the adaptive versus non-adaptive treatment revealed a marginally significant difference for players 15 years and older (adaptive: $M = 8.00$, $SE = .40$; non-adaptive: $M = 7.20$, $SE = .36$; $p = 0.08$, $d = .88$), see Figure 2.

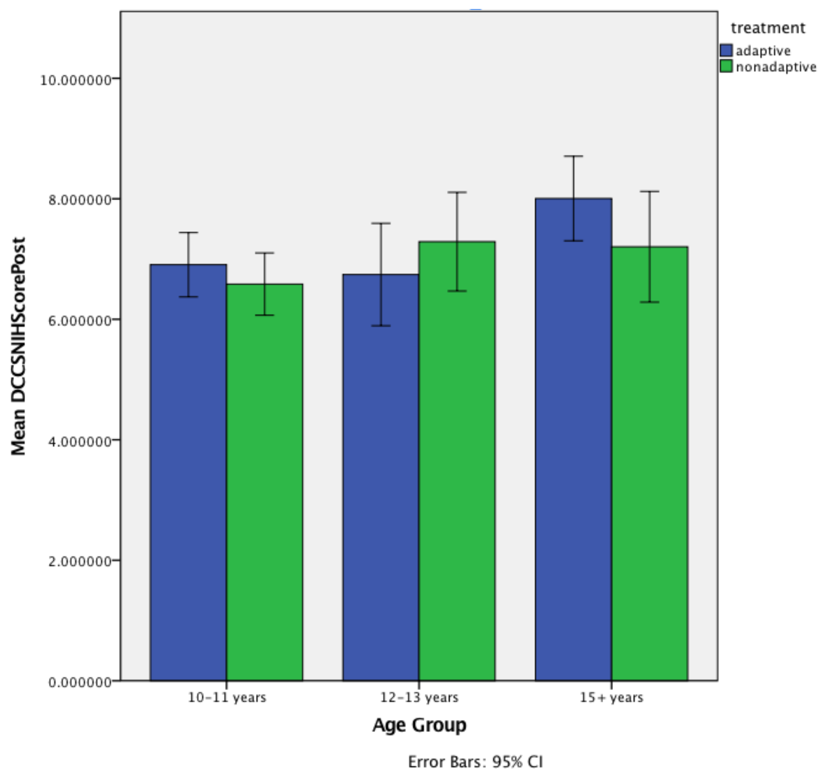

Figure 2. Groups differences for Adaptive versus non-adaptive conditions

To investigate to what extent the adaptive algorithm was able to support the gameplay experience of different age groups, we analyzed game log data of participants in the adaptive condition group. For this analysis, we included the 44 participants from the adaptive treatment group (45% Female). The final breakdown of this sample by age group is provided in Table 4.

As described above, in the adaptive version of the game, the game difficulty depended on participants' performance: for every 3 correct responses, the aliens' speed would increase, and for every incorrect response, the aliens would slow down. The speed of each alien was determined logarithmically, such that changes at the highest and lowest speeds were less dramatic than changes in the middle of the range. To study the differences in gameplay due to this feature, we analyzed three variables of interest: **count of adaptive changes, relative speed of aliens, and ratio of difficulty-increasing changes to all trials.**

Adaptive trials. During gameplay, we logged each instance where the difficulty increased, decreased or stayed constant. The distribution of increasing, decreasing, or same-speed trials differed significantly by age group, X^2 (4) = 66.44. $p < .001$, indicating that age groups required different patterns of adaptivity. This supports the notion that the adaptive algorithm was able to generate different levels of game difficulties for the different age group in order to keep them in their Zone of Optimal Engagement.

We then calculated an average count of adaptive trials for each participant to determine the influence of the adaptive adjustments on a participant’s gameplay experience. Results showed that there were differences in counts between age groups. The 15+ age group had the highest average number of trials for which difficulty increased, followed by the 10-11 year group and the 12-13 year group respectively (see table 3). Post-hoc Tukey tests revealed the difference between the 15+ group and the 10-11 group was significant, $p < .008$, as was the difference between the 15+ group and the 12-13 group, $p < .001$. The 10-11 and 12-13 year groups did not differ significantly, $p = .31$. The 10-11 age group had the highest average number of trials for which difficulty decreased, although this difference was not significant (post hoc Tukey tests p’s $> .14$). These findings further support the notion that the adaptive engine was able to keep players in their Zone of Optimal Engagement, which meant that for older learners, the difficulty was increased more often than for the other ages, and for the younger learners, difficulty was decreased more than for the other age groups.

Relative speed of aliens. During each difficulty adjustment done by the adaptive engine, the scale of change to the alien speed was logged. The range of these changes was between -15 to 10, with -15 being the lowest and 10 the highest, relative to the default speed of 1 in the non-adaptive version of the game. For example, a relative speed of 5 meant that aliens speed was 4 scale points above the default, and a relative speed of -3 meant that alien speed was 4 scale points lower than default. For this outcome, there was a significant effect of age group $F(2, 3132) = 84.73$, $p < .001$, $\eta_p^2 = .051$. We also observed a large difference between the 15+ year group and the other two age groups. Post hoc Tukey tests revealed the 15+ group was significantly faster than both the 10-11 year group ($t = -12.122$, $p < .001$) and the 12-13 group ($t = -11.40$, $p < .001$), although the difference between the younger groups was not significant ($t = 0.372$, $p = .926$). The results suggested that the 15+ year group was better able to react to faster aliens with the correct response, compared to the other groups.

Table 3: Means of game log statistics by age group

Age group	N	Increased Trials (SD)	Decreased Trials (SD)	Total Trials (SD)	Relative alien speed (SD)	Increase ratio (SD)
10-11	19	43.052 (12.007)	37.789 (16.130)	266.89 (32.971)	-0.006 (3.224)	0.159 (.031)
12-13	13	37.076 (11.094)	34.923 (17.188)	242.38 (41.204)	-0.056 (3.222)	0.152 (.030)

Age group	**N**	**Increased Trials (SD)**	**Decreased Trials (SD)**	**Total Trials (SD)**	**Relative alien speed (SD)**	**Increase ratio (SD)**
15+	8	58.125 (9.015)	24.750 (13.134)	289.75 (17.894)	1.825 (3.355)	0.200 (.025)

Ratio of difficulty-increasing changes. We calculated a difficulty increase ratio for participants by dividing the total number of increase changes with the total number of all changes. This ratio sheds light on the directionality and magnitude of adaptive changes. Further, the ration it indicates how difficult an adaptive game session was compared to the non-adaptive version (more increases in difficulty lead to a more difficult gameplay experience). The range of this ratio is between 0 to 1, with 0 being the least difficult and 1 being the most difficult. A summary of mean ratios by age group is provided in table 3.

Results of a one-way ANOVA showed that the differences in difficulty increase ratio between age groups were significant, $F(2, 37) = 7.12$, $p = .002$, $\eta_p^2 = .278$. Post hoc Tukey tests revealed that both the 10-11-year-old group and the 12-13-year-old group had a significantly lower proportion of increase trials than the 15+ group ($t = -3.25$, $p < .05$; $t = -3.58$, $p < .05$) respectively. The difference between the 10-11 and 12-13 groups was not significant, $t = 0.67$, $p = .78$. This indicates that the 15+ group had more correct responses as a proportion of their games, and thus a more difficult session of gameplay, than both the 10-11 and 12-13 groups.

Gameplay and executive function. We also explored the participants' gameplay data for associations with the executive function outcomes. For participants in the adaptive condition, we examined correlations between pretest and posttest DCCS scores with the percent of correct responses, Increase Ratio, and Decrease Ratio. For participants in the nonadaptive condition, pretest and posttest DCCS scores were correlated with only the percent of correct responses as an indication of game performance

Results revealed that percent of correct and DCCS posttest were significantly correlated for both the Adaptive and Nonadaptive groups. While pretest score was significantly correlated with performance in the adaptive condition, it was not significantly correlated with performance in the nonadaptive condition.

Table 4: Pearson Correlations Between EF Measures and Gameplay

	DCCS Pretest	**DCCS Posttest**	**Percent Correct**	**IRatio**	**DRatio**
Adaptive					
DCCS Pretest	—				
DCCS Posttest	0.28	—			
Percent Correct	0.47**	0.45**	—		
IRatio	0.32	0.38*	0.88***	—	
DRatio	-0.25	-0.55***	-0.85***	-0.72***	—
Non Adaptive					
DCCS Pretest	—			—	—
DCCS Posttest	0.16	—		—	—
Percent Correct	0.18	0.60***	—	—	—

* $p < .05$ ** $p < .01$ *** $p < .001$

Discussion

Experiment one revealed a significant general improvement in EF scores from DCCS pre- to posttest, which suggests that all participants benefited from playing the game. These results support previous evidence showing improvement in EF with game-based-training (Homer et al., 2018; Parong et al., 2017). The results also showed that the adaptive treatment did not result in significant EF gains over the non-adaptive control group. This finding suggested that both the adaptive and the non-adaptive versions of the game were equally effective at improving EF outcomes for students aged 10-15+. Since this age range may be too broad to expect a general effect, additional analyses involving age as variable were warranted.

Further analysis revealed a significant difference by age: Players 15 years and older performed better on the DCCS posttest than both younger groups of players. Exploratory analyses also revealed a trend showing that the effect of adaptivity on EF gains may be moderated by age. Players in the 15+ years groups benefited from the adaptive difficulty adjustment (d = .88), whereas younger players in the 10-11 and 12-13-year-old group did not.

The difference in outcomes between age groups reflects expectations related to EF development – younger players are still developing certain aspects of their EF, resulting in lower overall performance, whereas older players have more developed EF skills, especially in relation to the switching skill investigated in this research, resulting in higher overall performance.

The difference for the players 15 and older observed indicate that the adaptive engine may result in increased outcomes compared to the non-adaptive game for this age group, whereas no differences were found for the younger players. A lack of significant effect of adaptivity for the younger age groups indicates that the adaptive and the non-adaptive version were equally effective for them, suggesting that the difficulty of the non-adaptive version was appropriate for the younger learners. This is not surprising, as the original, non-adaptive levels of the game were designed to benefit adolescents of this age range. However, the non-adaptive version of the game appeared to be too easy for the 15+ age group, which benefited more from the adaptive adjustments in difficulty. Because the interaction of age and treatment failed to reach significant at the .05 level ($p = .08$), caution should be taken in assuming an age by condition interaction.

The user logs generated by the game provided insights in the functioning of the adaptive algorithm of the game. First, we found that the effect of adaptivity on gameplay was different between age groups. The count of adaptive changes suggest that the adaptive engine was making more adjustments for the 10-11 year group and 15+ year group, compared to the 12-13 year olds. Second, adaptivity played different roles for different age groups. For participants 15 and older, the adaptive engine provided an opportunity to play at a higher level of challenge than the non-adaptive version. This means that older players, who were on average performing well, were challenged more by the adaptive version of the game than by the non-adaptive version, leading to a higher chance for improvement. Using the lens of the Zone of Optimal Engagement, the non-adaptive version was at an appropriate difficulty to keep the younger groups in their Zone of Optimal Engagement but was too easy for the 15+ age group. Hence, the adaptive version had a significant effect on the 15+ age group as it increased the difficulty, giving them challenges at their skill level, i.e., within their Zone of Optimal Engagement. This pattern is supported by the Increase ratio statistics showing a higher ratio of difficulty increase changes for 15+ group than the younger groups. Third, the adaptive engine created customized gameplay experiences based on player performance. The Increase ratio and alien speed data show that groups had unique experiences with varying difficulty arcs, and that these arcs were in line with expected developmental differences in EF for these age groups.

Discussion and Conclusion

The goal of the research presented here was to determine whether the effectiveness of a game specifically designed to train executive function skills can be enhanced by adding a feature that was able to adapt the difficulty level of the training tasks to the current abilities of a particular individual, and specific age groups, and thereby keep players in the Zone of Optimal Engagement. We were especially interested in the question whether such adaptivity is equally effective for learners of different ages.

The game used in this research, the *Alien Game*, was designed specifically to help individuals increase their EF skills of shifting. The game was playtested with a large number of 12-16-year-old learners. Research provided empirical evidence that playing the game resulted in increased EF after approximately 1.5 hours of play for younger children (Homer et al., 2018) and after approximately 2 hours of play for college students (Parong et al., 2017). These numbers are much lower than what has been reported for other EF games in the literature and allowed us to run the experiments presented here with relatively short exposure times.

This experiment found that 20 minutes of game play resulted in a statistically significant gain of EF skills for all participants, independent of the version of the game they received. This result should not be interpreted as a claim that playing this game for only 15-20 minutes results in lasting changes in EF skills, but it does suggest that the tasks performed in the game make demands of the same cognitive resources related to shifting as the DCCS, and that a longer treatment could be able to result in longer term changes in players' EF.

We examined whether the adaptive game was more effective than the non-adaptive game for middle- and high school students, and found an effect for players 15 years and older. There was a trend suggesting the adaptive version was more effective than the non-adaptive version. This finding can be interpreted in different ways. One observation is that the game was designed and playtested with younger learners, which makes it less surprising that for these learners, no adaptivity was required to keep learners in the Zone of Optimal Engagement.

Limitations

Some of the design choices made in this research may limit the generalizability of our findings. First, 20 minutes of gameplay is brief, and we cannot speak to how long these effects may last. We are in the process of replicating these findings with longer interventions. The significance of the present study is that it demonstrates that there were observable differences that warrant further research on adaptivity. Second, the reported study only examines near-transfer effects of EF training. Near-transfer refers to studying transfer of EF training to distinct-yet similar tasks, while far-transfer is associated with improvement on cognitively unrelated tasks after EF training. Although studying near-transfer effects is the first crucial step in uncovering the link between EF training and far-transfer, it only partially uncovers the causal link between EF and success at relevant variables like math, reading, and other academic skills. Third, this experiment only explores the training of shifting sub-skill of EF. Hence, the findings may not generalize to training of other EF skills. However, because our adaptive engine is based on a broader framework of cognitive training (Lövdén et al., 2010), findings from this study may still hold for other cognitive training interventions.

limitation is the type of task difficulty in the adaptive condition. In the fixed condition, difficulty increases between levels due to the level of intersecting detail of rules for feeding the aliens, while in the adaptive condition, difficulty also increases within levels due to speed and the pressure of delivering correct responses in time. The additional challenge may be due to perceptual and motor skills, rather than cognitive switching skills, and may advantage participants who are gamers.

Conclusion

These experiments have important theoretical as well as practical implications. On the theoretical side, these studies showed that adaptive games can keep the learner in their Zone of Optimal Engagement in which the intervention is most effective. Our findings also reflect the developmental differences in EF for players of different ages, where older adolescents have more developed EF than younger ones. On the practical side, results provide an example of the importance of investigating design factors in game-based training of executive function. Although significant differences were not found between the two treatments, both versions were shown to train EF significantly. Analysis of user logs show that the game was in fact able to adjust to the specific learners' needs. More research is needed, but our game was able to show significant increases in EF even after 20 minutes of play, and suggested that using adaptivity may be an effective way to train executive functions, especially for older adolescents.

Acknowledgements

This work was partially supported by IES Grant R305A150417, Focused Computer Games that Promote Executive Functions Skills, and by PSC-CUNY Award #60816-00 48: Virtual Reality and Emotional Design for the Development of Executive Functions in Adolescents.

References

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, *58*(4), 375-404.

Bainbridge, K. & Mayer, R.E. (2017). Shining the Light of Research on Lumosity. *Journal of Cognitive Enhancement*, 1–20.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychological Bulletin, 121(1),* 65.

Best The News. (2016). Value of the global video games market from 2011 to 2020 (in billion U.S. dollars). In Statista - The Statistics Portal. Retrieved May 14, 2018, from https://www.statista.com/statistics/246888/value-of-the-global-video-game-market/.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647-663.

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.

Bryce, D., Whitebread, D., & Szűcs, D. (2014). The relationships among executive functions, metacognitive skills and educational achievement in 5 and 7 year-old children. *Metacognition and Learning*. doi:10.1007/s11409-014-9120-4

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science*, *6*(1), 3-5.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? *Contributions of inhibitory control and working memory. Infant and Child Development*, 11(2), 73–92. doi:10.1002/icd.298

Carlson, S.M., Zelazo, P.D., Faja, S. (2013). Executive function. In P.D. Zelazo (Ed.), *Oxford Handbook of Developmental Psychology, Vol. 1: Body and mind*, pp. 706-743. Oxford University Press, New York.

Charlton, B., Williams, R. L., & McLaughlin, T. F. (2005). Educational Games: a technique to accelerate the acquisition of reading skills of children with learning disabilities. *International Journal of Special Education, 20(2)*, 66-72.

Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media, 29*(3), 241-250.

CREATE (1995). *Alien Game*. CREATE Lab, New York University. Available create.nyu.edu/dream.

CREATE (2016). *Gwakkamole*. CREATE Lab, New York University. Available create.nyu.edu/dream.

CREATE (1996). *Digital Reference, Experiment and Assessment Manager DREAM*. CREATE Lab, New York University. Available create.nyu.edu/dream.

Conchas, G. (2006). *The color of success: Race and high achieving urban youth*. New York: Teachers College Press.

Desimone, L. (1999). Linking parent involvement with student achievement: Do race and income matter. The *Journal of Educational Research, 93*(1), 11–30.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science, 333*(6045), 959-964.

Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology, 50(6),* 1698.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental science, 12*(4).

Homer, B.D., Plass, J.L., Rafaele, C., Ober, T. M., & Ali, A. (2018). Improving high school students' executive functions through digital game play. *Computers and Education, 117*, 50-58.

Hromek, R., & Roffey, S. (2009). Promoting Social and Emotional Learning with Games: "It's Fun and We Learn Things". *Simulation & Gaming, 40(5),* 626-644.

Karbach, J., & Verhaeghen, P. (2014). Making working memory work: a meta-analysis of executive-control and working memory training in older adults. *Psychological science,* 25(11), 2027-2037.

Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay? *Computers and Education, 51(4),* 1609-1620.

Klingberg T, Fernell E, Olesen P, Johnson M, Gustafsson P, Dahlstrom K, Gillberg CG, Forssberg H, Westerberg H. (2005) Computerized training of working memory in children with ADHD – A randomized, controlled, trial. J*ournal of American Academy of Child and Adolescent Psychiatry, 44*(2), 177-186.

Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychological bulletin, 136*(4), 659.

Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology*, *50*(6), 1698.

Mayer, R. E. (2014). *Computer games for learning: An evidence-based approach*. MIT Press.

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.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current directions in psychological science, 21*(1), 8-14.

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review, 18*(1), 46-60.

Müller, U., & Kerns, K. (2015). The Development of Executive Function. *Handbook of Child Psychology and Developmental Science, Cognitive Processes, 2*, 571.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception.
Cognitive Psychology, 9, 353–383.

Nouchi, R., Taki, Y., Takeuchi, H., Hashizume, H., Akitsuki, Y., Shigemune, Y., & Kawashima, R. (2012). Brain training game improves executive functions and processing speed in the elderly: a randomized controlled trial. *PloS One, 7(1)*, e29676.

Pajitnov, A. (1989). Tetris [Nintendo DS]. Kyoto, Japan: Nintendo

Parong, J., Mayer, R.E., Fiorella, L., MacNamara, A., Homer, B.D., & Plass, J.L. (2017). Learning executive function skills by playing focused video games. *Contemporary Educational Psychology, 51*, 141–151. Doi: 10.1016/j.cedpsych.2017.07.002

Powers, K. L., Brooks, P. J., Aldrich, N. J., Palladino, M. A., & Alfieri, L. (2013). Effects of video-game play on information processing: a meta-analytic investigation. *Psychonomic Bulletin & Review, 20*(6), 1055-1079.

Plass, J.L. (2016). A Taxonomy of Adaptivity in Learning. Panel on *Personalized and Adaptive Learning Systems, CRESSTCON '16*. September 20, 2016 in Los Angeles, CA.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist, 50*(4), 258-283.

Riggs, N. R., Jahromi, L. B., Razza, R. P., Dillworth-Bart, J. E., & Mueller, U. (2006). Executive function and the promotion of social–emotional competence. *Journal of Applied Developmental Psychology, 27(4)*, 300-309.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.

Qu, L., Wang N., & Johnson, W. L. (2005). Detecting the learner's motivational states in an interactive learning environment. In C.-K. Looi et al. (Eds.), *Artificial Intelligence in Education* (pp. 547-554), IOS Press.

Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research, 64*(1), 37-54.

Van Muijden, J., Band, G. P., & Hommel, B. (2012). Online games training aging brains: limited transfer to cognitive control functions. *Frontiers in Human Neuroscienc*e, 6, 221.

Von Bastian, C. C., & Eschen, A. (2016). Does working memory training have to be adaptive? *Psychological Research*, *80*(2), 181-194.

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors*, 5, 7–17.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature protocols*, 1(1), 297.

Zelazo, P. D., & Bauer, P. J. (Eds.). (2013). National Institutes of Health Toolbox cognition battery (NIH Toolbox CB): Validation for children between 3 and 15 years. Wiley.