Title
Cognitive Consequences of Short and Long Term Training with Lumosity Games

Author
Bainbridge, Katie

Publication Date
2017

UNIVERSITY OF CALIFORNIA

Santa Barbara

Cognitive Consequences of Short and Long Term Training with Lumosity Games

A Thesis submitted in partial satisfaction of the
requirements for the degree Master of Arts
in Psychological and Brain Sciences

by

Katie Bainbridge

Committee in charge:

Professor Richard Mayer, Chair

Professor Mary Hegarty

Professor David Sherman

Professor Jonathan Schooler

June 2019

The thesis of Katie Bainbridge is approved.

Dr. Jonathan Schooler

Dr. David Sherman

Dr. Mary Hegarty

Dr. Richard Mayer, Committee Chair

June 2019

ABSTRACT

Cognitive Consequences of Short and Long Term Training with Lumosity Games

by Katie Bainbridge

Lumosity is a subscription-based suite of online brain-training games, intended to improve cognitive skills. Due to an influx of products designed to train cognition through games such as Lumosity, it is important to determine their effectiveness for the sake of consumers and for the potential implications of any training effects for theories of transfer of cognitive skills. Two training experiments were conducted using the Lumosity platform. Participants were divided into three groups: those who trained with five attention games in Lumosity (attention group), those who trained with five flexibility games in Lumosity, (flexibility group), and an inactive control group. Participants were assessed on accuracy and response time for two cognitive tests of attention (Useful Field of View and Change Detection) and two cognitive tests of flexibility (Wisconsin Card Sort and Stroop) both before and after a training period. In Experiment 1, the training period was 3 hours spread over 4 sessions. In Experiment 2, the training period was 15 to 20 hours spread over 80 sessions. The trained groups did not show significantly greater pretest-to-posttest gains than the control group on any measures in either experiment, except in Experiment 2 the flexibility group significantly outperformed the other two groups on Stroop response time, which is very similar to one of the flexibility games. A practical implication concerns the lack of evidence for the effectiveness of brain training games to improve cognitive skills. A theoretical implication concerns the domain-specificity of cognitive skill learning from brain training games.

TABLE OF CONTENTS

Shining the Light of Research on Lumosity

Objective and Rationale

Visionaries foresee a future in which people can get smarter simply by playing appropriately designed computer games (Gee, 2007; Prensky, 2006; McMonigal, 2011; Shaffer, 2006; Squire, 2011), but research evidence to support these speculations is lacking (Honey & Hilton, 2011; Mayer, 2014; O'Neil & Perez, 2008; Tobias & Fletcher, 2011; Wouters & van Oostendorp, 2017). The present study addresses this gap between strong claims and weak evidence by conducting well-controlled experiments examining the effects of playing well-designed computer games on improvements in the cognitive skills targeted by the games.

Brain training games--such as the classic Nintendo game, *Brain Age*--represent a genre of computer games intended to train specific cognitive skills or to improve cognitive functioning in general (Mayer, 2014). Our focus in the present study is on the effectiveness of *Lumosity*, a suite of brain training games intended to improve the player's cognitive skill in a number of areas such as attention--the ability to visually track a target in a complex visual field--and flexibility--the ability to shift rapidly from one task to another. For example, Figure 1 shows a frame from a *Lumosity* game intended to improve attention skill--*Eagle Eye*, in which the player is required to rapidly identify a briefly appearing target among a field of distractors. Similarly, Figure 8, shows a frame from a *Lumosity* game intended to improve flexibility skill--*Color Match*, in which the player is required to indicate whether the meaning of the word on the left matches the color of the word on the right. *Lumosity* is sold online on a subscription basis, with the motto: "Enjoy brain training created by scientists and game designers." (lumosity.com, retrieved October 25, 2016). The goal of the present study is to provide evidence concerning the effectiveness of cognitive training with Lumosity games both over a short term (3 hours over 4 sessions) and long term (15 to 20 hours over 80 sessions).

Rationale

Whether brain-training games, such as *Lumosity*, can help people improve their cognitive skills is an important practical and theoretical question, particularly in regards to adding to our understanding of how and if cognitive skills can improve. On the practical side, computer games designed to improve cognitive skills are being sold to the public (Redick, 2013), so it is worthwhile to help consumers make informed choices. On the theoretical side, research on the cognitive consequences of playing brain training games has implications for theories of transfer of cognitive skill, particularly concerning the degree to which skills exercised in a game transfer to tasks outside the game (Mayer, 2014).

The argument concerning the effectiveness of brain training games has been contentious both in the legal arena and the scientific arena. In the legal arena, in 2016, Lumos Labs, the company that sells *Lumosity* agreed to settle a suit for false advertising brought by the Federal Trade Commission, without admitting or denying the allegations (Robbennolt, 2016). As part of the settlement the company agreed to stop making claims about Lumosity's effects on performance or cognitive impairment without supporting scientific evidence. In the scientific arena, in 2014, the lack of evidence for the effectiveness brain training games was addressed by the scientific community in a letter signed by nearly 70 researchers in 2014, but was soon countered by a letter signed by 133 supporters of brain training games (Simons et al., 2016). A recent review concluded that there is strong evidence that brain training games can improve performance on the trained tasks but there is not strong evidence that learning transfers to improvement in cognitive skills performed outside the game environment (Simons et al., 2016). The authors criticized the existing research base and called for well-controlled experiments that clearly examine the effects of playing brain training games. We take up that call in the present set of experiments.

Cognitive Consequences Approach

Cognitive consequences research is an ideal methodology for examining this question, as described by Mayer (2014, p. 172): "The cognitive consequences approach to game research compares the cognitive skill performance of students who are assigned to play an off-the-shelf computer game for an extended period (game group) to those who engage in an alternative activity (control group)." (Mayer, 2014, p. 172). Based on the cognitive consequences approach, we are comparing pre-to-posttest gains between groups in order to determine whether brain training games improve performance on cognitive tasks outside of the game.

There are three possible outcomes concerning the effect playing brain training games will have on cognitive skills (Mayer, 2014):

1. Playing a game will only improve skills directly related to the game. A player will improve on the game, but that improvement will not transfer to any other situation. If you were to learn chess, the skills you acquire will only apply to chess. If you learn to play chess in a park and then discover you can play chess on your computer, your prior knowledge of chess will benefit you, but that prior knowledge will not meaningfully benefit you in any other context. We refer to this as *specific transfer*.
2. Playing a game will engage and strengthen targeted cognitive skills, and players will be able to apply these improvements to new, but related situations. If you were to learn chess, the skills that chess challenges, such as keeping multiple outcomes in mind, will be strengthened. When a new environment requires you to consider a

variety of options simultaneously, your performance will benefit from the practice you gained while playing chess. We refer to this as *specific transfer of general skills*. This is the marketed effect of commercial cognitive training games, i.e. playing a memory game will improve your memory.

3. Playing a game will train the mind in general and as a result a wide variety of cognitive skills will improve. If you were to learn to play chess, your intelligence will increase overall. Perhaps you will be able to plan competent battle strategies or learn a language more easily. The more you master the game, the better you will perform in a variety of cognitive realms. We refer to this as *general transfer*.

The goal of the present studies is to determine which of these outcomes current cognitive training games can engender. In particular, this research is intended to address theoretical issues concerning the breadth of transfer for cognitive training, and to address practical issues that could help inform consumers (Shipstead, 2012) and game developers by assessing which aspects of games are effective and where improvements could be made.

Game Training Platform

Lumosity was chosen as the subject for our cognitive consequences research. It is from a company that offers brain-training games through a web interface. This platform was chosen in part because the company classifies their games by domain, consistent with the specific transfer of general skills theory of cognitive training. *Lumosity* is also quite popular and claims to be research-based (Hardy, 2011). It was also chosen because it offers a multitude of games (8-14) in each domain, and previous research has suggested that too much repetition encourages specific transfer; variety is necessary if improvements are going to transfer to new situations (Green & Bavelier, 2012). The games are adaptive, so players are constantly being challenged at the upper bound of their abilities in a given game. The combination of these factors provides reason to believe that *Lumosity* games could improve skills in novel applications of the trained domain, as their design suggests.

Does Lumosity Work?

In reviewing the literature on the effectiveness of *Lumosity* training, we employed the following criteria for inclusion: (a) random assignment of participants, (b) inclusion of a control group, (c) healthy adult population, (d) training of sufficient length using games designed to improve cognition on a computer interface, and (e) published in a peer-reviewed and ISI-indexed journal. The existing research literature concerning the efficacy of *Lumosity* can be divided into studies sanctioned by *Lumosity's* parent company, Lumos Labs, and those by third-party researchers. Ultimately our criteria had to be relaxed in order to include *Lumosity* studies in the review.

Research from Lumosity. Lumosity lists a number of research studies on their website as of August, 2015. Of the 13 studies listed, 8 are posters that have not yet been published, 5 do not investigate efficacy (some of these studies overlap), 1 is published in a non-indexed journal, and 1 has no control group. Neither of the 2 remaining studies includes a healthy, normal population, but they are included in this review nonetheless.

One of these studies trained executive function in 41 women receiving chemotherapy for breast cancer, in 48 sessions spread over 12 weeks using *Lumosity* games (Kesler, Hosseini, Heckler, Janelsins, Palesh, et.al, 2013). Twenty of these participants served as the inactive control group in the form of a waitlist that eventually received training. The *Lumosity* games chosen were intended to train cognitive flexibility, working memory, processing speed, and verbal fluency. After the training period, participants showed significant improvements on 3 out of 7 post-tests, specifically the Wisconsin Card Sorting Task (d = 0.74), letter fluency (d = 0.39), and a symbol search task (d =1.00). Although this study produced some significant training effects, impaired cognition is a side effect of chemotherapy and it is not clear whether the effects would apply to a healthy population.

The other study of suitable design recruited 16 older adults with mild cognitive impairment to complete 30 training sessions consisting of 6 Lumosity games (Finn & McDonald, 2011). Eight of these participants were put on a waitlist to act as the inactive control group. These games were designed to train attention, processing speed, visual memory, and cognitive control. Before and after the training period, the participants were assessed on set shifting, visual learning, visual recognition, visual sustained attention, and visual working memory using the CANTAB neuropsychological evaluation. Of these, the only significant outcome after treatment was for visual sustained attention ($d = 1.17$). Correlations were conducted between game score performance and post-test performance, and while many of the domains showed significant correlations between training and post test performance, visual sustained attention post test scores (i.e. the only significant outcome) were not correlated with game performance. Although *Lumosity* promotes the study on their website as proof of their product's success, it should be noted that no significant effects were found for most of the cognitive skill measures and it is unclear if the one positive effect would apply to healthy subjects.

In summary, of the 13 studies listed on the *Lumosity* website, only two meet the minimal standards of a randomized, controlled study published in a peer-reviewed journal that trains with a product designed to improve cognition. Of those, neither found significant effects for most of the skills measured and neither involved healthy adult subjects. Thus, we conclude the research highlighted by *Lumosity* does not provide convincing evidence that playing *Lumosity* games improves cognition in healthy adults. In fact, none of the study designs was capable of providing evidence that *Lumosity* is effective in healthy adults.

Evidence from third party sources. A study from Redick, Shipstead, Harrison, Hicks, Fried, and Engle et al. (2013) recruited 75 participants from multiple colleges of various levels of prestige. They divided them into a group that trained with an adaptive n-back working memory task for 20 sessions and an active control group that practiced an adaptive visual search task. Visual search skill had previously been shown to be unrelated to working memory capability (Kane, Poole, Tuholski, & Engle, 2006), making it an appropriate active control condition. The researchers tested the participants at three separate intervals (0 sessions, after 10 sessions, after 20 sessions) with 17 transfer tests to assess whether the amount of training would influence the amount of cognitive gain. They found significant improvements on the tasks trained (the n-back and the visual search task) but no evidence of transfer to any of the 17 tests assessing intelligence, multitasking, working memory capacity, or perceptual speed, and no dose-response effect (Redick et al, 2012).

Increased sample size does not seem to make elusive transfer effects more apparent. Owen, Hampshire, Grann, Stenton, Dajani, Burns, et al. (2010) recruited 11,430 people to train with a series of online tasks. The first group practiced tasks designed to improve reasoning, planning, and problem solving. The second group practiced tasks designed to improve memory, attention, visuospatial processing, and mathematics. The control group answered obscure questions using the Internet. Participants were expected to complete six training tasks for 10 minutes a day, three times a week for six weeks, totaling in 3 hours of training. They were assessed using a common neuropsychological battery of four tests chosen for its known sensitivity to minute neuropharmacological changes in healthy adults. All three groups improved comparably, suggesting a test-retest practice effect and nothing more (Owen et al, 2010). The participants trained for only about 3 hours, so it is possible the training period was not sufficient.

Meta-analyses of other brain training platforms paint a similar picture. A review was done of all papers that had conducted adaptive working memory training using CogMed for at least two weeks. There was no evidence that CogMed training improved performance on any metric either in typical populations or in participants with working memory dysfunction (Hulme & Melby-Lervag, 2012). Admittedly, CogMed is a different platform than *Lumosity*, and perhaps another product would be more effective.

The same authors conducted another meta-analysis of any well-controlled and randomized working memory training study regardless of platform. Of the 23 studies included, the analysis found at best short-term, near-transfer effects. The studies did not reveal any far transfer effects and no effects at follow-up. Even the short-term, near-transfer effects are dubious, as they found that "in the best designed studies, using a random allocation of participants and treated controls, even the immediate effects of training were essentially zero" (Melby-Lervag & Hulme, 2012, p. 281).

The only suitably well-designed brain training study that features *Lumosity* rather than another game company comes from Shute, Ventura, and Key (2015). Seventy-seven Florida State University undergraduates were randomly assigned to play either the popular action video game *Portal 2* (n = 44) or *Lumosity* (n = 35) for 8 hours over the course of 1-2 weeks. A cognitive battery assessing problem solving, spatial skills, and persistence was administered both before and after the study. After 8 hours of game training, the participants who played *Portal 2* improved significantly more than *Lumosity* players in all categories: problem solving ($p = .02$, $d = .59$), spatial skills ($p = .04$, $d = .64$), and persistence ($p = .05$, $d = .42$). No significant improvements were seen in those who trained with *Lumosity* for 8 hours (Shute, Ventura & Ke, 2015). In this study both groups were active conditions, so this study does not speak to whether *Lumosity* improves performance better than an inactive control. In addition, their *Lumosity* condition played all 52 games available in the *Lumosity* suite, not just games associated with the researchers' cognitive measures of problem solving and spatial skills. Perhaps more targeted skill training, in keeping with the specific transfer of general skills theory, would yield better results.

In a large-scale on-line controlled trial, Hardy et al. (2015) found that people who were assigned to a program of *Lumosity* games for 50 15-min sessions showed greater pretest-to-posttest gains on a variety of cognitive tests as compared to an active control group that solved crossword puzzles, with an average effect size of $d = 0.25$. Simons et al. (2016, p. 143) state that "these significant effects should be interpreted with caution" in light of methodological flaws and potential conflicts of interest in which five of the seven authors are employees of the company that sells *Lumosity*. Our goal in the present study was to conduct a well-controlled, long-term trial that would not be subject to any conflicts of interest.

How to Improve Cognitive Training Research

Overall, we conclude that there is not sufficient evidence to support or even test claims that *Lumosity* brain training games are effective at improving cognitive skills in healthy adults. Simons et al. (2016, p. 143) come to a similar conclusion concerning the available evidence concerning the effectiveness of *Lumosity* as a vehicle for improving cognitive skills: "In sum, Lumos Labs cites little, if any, compelling evidence from randomized controlled trials that supports the claim that practicing *Lumosity* tasks yield broad improvements in cognitive abilities. Moreover, the evidence the company does cite mostly consists of non-peer reviewed studies or studies that could not (by design) provide such evidence."

The present study seeks to help fill the gap between claims and evidence, by contributing to the research base on the effectiveness of *Lumosity* as a brain training program. In order to assess which of the three outcomes outlined earlier–specific transfer, general transfer, or specific transfer of general skills–is more likely, a stringent experimental design is necessary. In particular, a suitable study includes sufficient sample size, random assignment, experimental control, appropriate measures of transfer of cognitive skill, and an appropriately long training period using games designed to improve cognitive skills (i.e., brain-training games).

Concerning sample size, some studies suffer from a low number of participants, leading to egregiously high variability, as in the Finn and McDonald study (2011). Concerning measures, a common neuropsychology battery can include numerous nonsignificant categories, as the tasks are not necessarily related to the trained skill. According to the specific transfer of general skills view, it is reasonable to predict that training one's ability to inhibit an old rule in a card sorting task could transfer to one's ability to inhibit saying the color word rather than the text color in the Stroop task. It is not reasonable to predict that training inhibition will transfer to quantitative reasoning skills, or some other similarly distant domain. The effortful aspects of the assessments selected should utilize similar, but not identical, skills to those trained, and should be in a non-game context.

Concerning experimental control, comparing a game-playing group to an inactive control group makes it difficult to draw causal conclusions. An active control group is more appropriate to ensure a placebo effect cannot explain the group differences. However, for most users the alternative to joining the *Lumosity* brain-training program is not some other cognitive training method, it is their normal routine. In this case, both active and inactive control groups are informative in separate and complimentary ways: the inactive group establishes the differences between users and non-users, while the active group helps elucidate the level of transfer.

In the current studies, we attempt to address these parameters in at least one of the two experiments reported in this paper. In both studies there are two active groups that train in separate cognitive skills (attention and flexibility), and one inactive group that does not train at all. The groups that train serve as active controls for each other, and the untrained group serves as an inactive control. The cognitive tests used to assess participants before and after were picked because the skills they utilize are very similar to those trained by the games selected for each condition (i.e. quickly identifying a target among distractors), but the format and mechanisms are different from the game environment. Both the games and tasks are described below. If the specific transfer of general skills theory is correct, as *Lumosity* asserts, these tasks should be sufficiently close to the trained skills to allow transfer. The first study has participants train for 3 hours spread over 4 sessions across 2 weeks in the lab. The second study extends this training period, and has participants train for 20 hours spread over 80 sessions across 16 weeks from their home computers.

Predictions

Three possible outcomes for this design exist. All three outcomes are informative, but speak to different conclusions about the efficacy of *Lumosity* brain training and the ability of cognitive training to transfer.

Hypothesis 1. According to the specific transfer theory, no group will discernably improve more than any other group. The group that trains on games for attention will improve on those games, but will perform equally on all posttests to controls and to the group that trains on games for flexibility. Likewise, the group that plays flexibility games will improve on those games, but will perform equally to the other two groups in posttest measures. This outcome would indicate that *Lumosity* game training is ineffective, or at least does not transfer outside of the game context to the quite similar context of the cognitive tests. If this level of near-transfer does not exist, it is very unlikely *Lumosity* training will improve the user's skills outside of the games in any meaningful way.

Hypothesis 2. According to the specific transfer of general skills theory, the group that trains on games for attention skill will perform equally to controls on posttests of flexibility but will exceed controls on posttests of attention, whereas the group that trains on games for flexibility skill will perform equally to controls on posttests of attention but will exceed controls on posttests of flexibility. This outcome would indicate that Lumosity games accomplished the goal of their design. Gains are domain-specific, leaving no doubt that playing a game designed to improve your ability to mentally inhibit first responses will result in an increased ability to inhibit responses in a multitude of contexts. The two training groups act as active controls for one another, ensuring the improvements are due to the targeted training and not from familiarity with the format or expectations of cognitive improvement.

Hypothesis 3. According to the general transfer theory, the group that trains on games for attention will exceed the control group on posttests of both attention and flexibility skill, whereas the group that trains on games for flexibility will also exceed the control group on posttests of both attention and flexibility. Perhaps the most interesting outcome, this result would indicate that *Lumosity* is successfully training the participants, but not solely in the domain they target (e.g. attention or flexibility). They may be training some other, perhaps non-cognitive skill, such as confidence, resilience, or familiarity with the computer interface.

Experiment 1

Experiment 1 is a randomized controlled experiment aimed at determining the cognitive consequences of playing *Lumosity* games for a short period.

Method

Participants. The participants were recruited from the subject pool at university in the western United States and the surrounding community. Some participants were compensated with course credit, and some were compensated with $10 per hour of participation. Overall 72 people between the ages of 18 and 31 finished (mean age = 20.8, *SD* = 2.39). An additional 14 people started the study, but were excluded because they dropped out before the final session. Of the participants surveyed, 25 were men and 28 were women. No survey was administered to the first 19 participants, so gender data is only available for 53 subjects. Random assignment put 26 people in the attention group, 25 people in the flexibility group, and 21 people in the control group.

Materials. The materials consisted of four computer-based cognitive skill tests and two sets of brain-training games.

Attention tasks. Two of the cognitive tests were intended to assess visual attention skills: the *Useful Field of View* (UFOV) and *Change Detection* tasks.

Useful Field of View. The UFOV task was used to assess visual processing speed (Edwards, et.al., 2005). The UFOV requires the participant to focus on a fixation point in the middle of the screen. It then very briefly flashes a target shape along 1 of 8 axes at various distances from the fixation point, and the participant must indicate along which axis the target appeared with the number keypad. There are 76 trials in the task and the target is presented for 17ms. The measures used for analysis were accuracy of identification (recorded as percent correct) and reaction time.

Change Detection. This task was used to assess visual attention (Mueller & Piper, 2014). This task alternates between two seemingly identical images of an array of colored dots of various sizes. The participant must identify the one dot in the array that changes size or shade. The measure used for analysis was how quickly the participants could identify the change. Experiment 2 added participant accuracy (recorded as percent correct), but it was not included as a measure in the Experiment 1.

Flexibility tasks. Two of the cognitive tests were intended to assess flexibility and inhibition skills: The *Wisconsin Card Sorting Task* (WCST) and the *Stroop task*.

Wisconsin Card Sorting task. This task was used to assess flexibility/inhibition skill (Berg, 1948). The WCST displays four cards with different shapes of different colors in differing amounts (for example, one card might show two green triangles while another will show four blue circles). The participant is then dealt a card and asked to sort it into one of the four piles based on an invisible rule--sort by number, sort by color, or sort by shape. Every few turns, the invisible rule changes without notice, and the participant must inhibit their previous strategy in favor of a new one. The measure used for this analysis was the number of perseverative errors, or errors made after the rule change is known.

Stroop task. This task was also used to assess flexibility/inhibition skill (Golden, 1958). The Stroop showed the participant a color word written in a different color than the word's meaning (e.g. "Yellow" written in the color red). The participant must identify the color of the text, ignoring the meaning of the word. They must inhibit their initial response to read the text in favor of the more flexible task of identifying the color. The measures used for this analysis were rate of accuracy and reaction time.

Attention games. Players assigned to the attention group played five *Lumosity* games designed to promote attention skills and designated as “Attention” games on the *Lumosity* website: *Eagle Eye, Birdwatching, Observation Tower, Top Chimp,* and *Space Junk*.

Eagle Eye. The player focuses on a fixation point at the center of the screen. A target bird shape and distractor shapes are flashed briefly along with a number in the fixation point. The participant must identify where the bird was and what number was shown in the center. Figure 1 shows a screenshot from the game.

Birdwatching. This game is very similar to *Eagle Eye*. Again, the player focuses on a fixation point at the center of the screen while a target bird shape and distractor shapes flash in the periphery. The player must correctly locate the bird and ignore the distractors; however, this time they must identify a *letter* in the fixation point, and the letters in each trial collect to form a hangman-style word problem that the player can solve to increase their score. Figure 2 shows a screenshot from the game.

Observation Tower. The player is shown several numbers scattered across the screen, but they are quickly covered. The player must click on the covered numbers in order-smallest to largest. The numbers they get correct turn into blocks that build a tower, with the goal of building the highest tower possible. Figure 3 shows a screenshot from the game.

Top Chimp. The mechanics in this game are very similar to *Observation Tower.* The player is paired with a cartoon chimp competitor. They are shown how long the numbers will be visible and must bet on how many numbers they can correctly order in that time. The numbers are flashed and again they are covered and the player must click them in order from smallest to largest. They must succeed in enough bets to beat the chimp. Figure 4 shows a screenshot from the game.

Space Junk. The player sees a black background through the window of a space ship. The game quickly flashes a given amount of space debris in the window. The player must accurately identify how many pieces of debris they saw. In later trials the variety of debris increases and players must identify how much of each kind of junk they saw. Figure 5 shows a screenshot from the game.

Flexibility games. Players assigned to the flexibility group played five *Lumosity* games designed to promote flexibility skills and designated as “Flexibility” games on the *Lumosity* website: *Ebb & Flow, Disillusion (new), Color Match, Brain Shift,* and *Brain Shift Overdrive*.

Ebb & Flow. The player uses the arrow keys to indicate direction as leaves move up, down, left, or right across the screen. When the leaves are green the player uses the arrow keys to indicate where the leaves are pointing. When the leaves are yellow the player uses the arrow keys to indicate where the leaves are moving. The colors of the leaves switch without warning. Figure 6 shows a screenshot from the game.

Disillusion (new). The player is shown an array of puzzle-like pieces of various colors with various shapes on them (e.g. crosses or circles). The array has a number of open spots around the edges. The player is given a target piece to place into the array with a sorting rule: if the piece is oriented vertically, the player sorts by color. If the piece is horizontal, the player sorts by shape. Figure 7 shows a screenshot from the game.

Color Match. The player is shown two color words side to side written in various text colors. They are asked to indicate with the arrow keys if the meaning of the word on the left matches the color of the text on the right. Figure 8 shows a screenshot from the game.

Brain Shift. The player is shown two white boxes stacked on top of each other. In one box at a time, the player is shown a combination of 1 letter and 1 number (e.g. the top box will say A7). They must indicate whether this combination follows the rule for the box that contains it- the top box asks whether the numbers are even and the bottom box asks if the letters are vowels. Figure 9 in the appendix shows a screenshot from the game.

Brain Shift Overdrive. This game doubles the premise from *Brain Shift*. Instead of two boxes there are four, and now the player must indicate if the letters are consonants and if the numbers when the combinations appear in their respective boxes. Figure 10 shows a screenshot from the game.

Survey. Participants were given a questionnaire after completing their final evaluation. The survey asked about their age and gender. It also asked about their video game habits, including how often they play action video games, casual video games, and brain training games, as well as whether they consider themselves a “gamer”. The survey was conceived and added to the procedure after 19 participants had already finished the study. As such, analyses done with survey data reflect only 53 participants and should be interpreted with caution.

Apparatus. The games and tests were administered on a Dell computer system with a 20-in screen.

Procedure. Participants were randomly assigned to a group and tested individually in a research room in a university psychology lab. Upon arrival the participant was seated in a cubicle containing a computer work station. The experimenter passed out and explained the consent form for the participant to read and sign. After the consent form was signed, the experimenter administered each of the four pretests in the following order: UFOV, Change Detection, WCST, and Stroop. Altogether the assessments took about 30 minutes. The participants were then randomly assigned to the attention, flexibility, or control group. The attention group trained their rapid visual processing and attention skills with 5 games: *Eagle*

Eye, Birdwatching, Observation Tower, Top Chimp, and *Space Junk*. The flexibility group trained their inhibition and rule shifting skills with the other 5 games: *Ebb & Flow, Disillusion (new), Color Match, Brain Shift*, and *Brain Shift Overdrive*.

In the active conditions, the first session consisted of 30 minutes of assessment and 30 minutes of training with the *Lumosity* games assigned to their group. The second and third sessions consisted of an hour of training each with the same games as in the first session. The final session began with 30 minutes of training with the same games and concluded with repeating the cognitive tests from the first session. The control condition consisted of just two 30-minute sessions, one in which participants completed the pretests and one in which they completed the same tasks as posttests, with no training or game-playing in between sessions. After the post assessment was completed, all conditions received a survey inquiring about their gaming habits and were then given their compensation. We adhered to guidelines for ethical treatment of human subjects and obtained IRB approval.

Results

Scoring. The main dependent measures used for the UFOV were accuracy (number correct divided by total trials) and reaction time in milliseconds. For the Change Detection Task reaction time was the only dependent measure used. For the WCST the dependent measures were number of preservative errors (errors made after the new rule has been established) and reaction time in milliseconds. The dependent measures used for the Stroop Task were number of errors and reaction time in milliseconds. Gain scores were also computed by subtracting pretest scores from posttest scores on all measures. Analysis of variance was used to test for differences among the three groups (i.e., attention, flexibility, and control) on pretests. ANCOVAs were conducted on posttest scores using pretest scores as covariates.

UFOV. The top section of Table 1a shows the mean and SD for participants on the UFOV for the pretest and posttest, as well as the pretest-to-posttest gain score for each group. The groups did not differ significantly on UFOV pretest accuracy, F (2, 60) = 0.77, p = 0.47, so any differences seen on the posttest could be attributed to our training intervention. An ANCOVA comparing UFOV accuracy posttest scores with pretest score as a covariate revealed no significant difference in improvement by group, F (2, 65) = 0.32, p = 0.73, η^2 = 0.01. Neither of the groups that trained with *Lumosity* games improved significantly more than the control group.

The same series of analyses was conducted with participants' reaction time on the UFOV, yielding the same pattern of results. The groups did not differ significantly on pretest mean response time on the UFOV, F (2, 69) = 0.12, p = 0.89, showing the groups were not different before they started. The groups also did not differ significantly on an analysis of covariance comparing groups on posttest reaction time using pretest as a covariate, F (2, 65) = 0.094, p = 0.91, η2 = 0.003, indicating that no group improved significantly more than any other. There is no evidence that playing 3 hours of *Lumosity* games focused on attention or flexibility skill improved performance on UFOV reaction time scores beyond a control group. The results for accuracy and response time on the UFOV task are consistent only with specific transfer theory (hypothesis 1), and do not support the idea that playing brain training games transfers to improvements in cognitive skills outside the game environment (hypotheses 2 and 3).

Change Detection. Reaction time was the only metric used to analyze the Change Detection task in Experiment 1. Means and SDs by group can be found in the bottom section of Table 1a. First an ANOVA was conducted on pre-test reaction time to see if the groups differed before the training. It was nonsignificant, *F* (2, 68) = 2.48, *p* = 0.09. An ANCOVA performed on posttest scores with pretest scores as a covariate revealed a significant difference by training group, F (2, 67) = 3.69, p = 0.03, η2 = 0.10. To elucidate the nature of this difference a Fisher's least significant difference (LSD) post hoc test was conducted. It revealed that the attention group improved significantly less than the flexibility and the control group, and the flexibility and control group did not differ significantly from each other. The positive ANCOVA result indicates the attention group's failure to improve rather than any group's superior gains. No group improved significantly more than the control group, which supports the specific transfer theory (hypothesis 1) but neither of the other theories (hypothesis 2 or 3).

WCST. The top section of Table 1b shows the means and SDs for WCST performance on the pretest and posttest as well as the pretest-to-posttest gain score for each group. The groups did not differ significantly on WCST pretest error scores, *F* (2, 69) = 0.10, *p* = 0.91, so any differences seen on the posttest can be attributed to the training intervention. An ANCOVA performed on posttest scores with pretest scores as a covariate revealed no significant differences among the groups, *F* (2, 68) = 0.37, *p* = 0.69, η2 = 0.01, indicating no group decreased in errors significantly more than any other. There is no evidence that playing 3 hours of *Lumosity* games focused on attention or flexibility skill improved performance on WCST beyond a control group.

The same series of analyses was conducted with WCST reaction time. Once again the pretest ANOVA was nonsignificant, *F* (2, 69) = 1.024, *p* = 0.346, indicating no significant difference among groups before training. An ANCOVA on posttest scores with pretest scores as a covariate revealed a significant difference among groups, *F* (2, 68) =

3.25, $p = 0.045$, $\eta2 = 0.09$. A post-hoc LSD showed that the flexibility group improved significantly more than the attention group, but neither the attention group nor the flexibility group improved significantly more than the control group. This once again suggests that the significant difference reflects the attention group's failure to improve, and does not provide evidence that training with Lumosity games will improve performance on WCST reaction time. This is consistent with the specific transfer theory (hypothesis 1) but not the other two hypotheses.

Stroop. The final section of Table 1b shows the means and standard deviations on the Stroop task pretest and posttest as well as the pretest-to-posttest gain scores for each group. The groups did not differ significantly on Stroop pretest error scores, F (2, 67) = 1.05, $p = 0.36$, so any differences seen on the posttest can be attributed to our training intervention. The groups did not differ significantly on posttest errors with pretest errors as a covariate, F (2, 63) = 0.08, $p = 0.92$, $\eta2 = 0.00$, indicating that no group decreased in errors significantly more than any other. There is no evidence that playing 3 hours of *Lumosity* games improved performance on Stroop accuracy more than a control group.

The same series of analyses was conducted with participants' reaction time on the Stroop task, yielding an identical pattern of results. The groups did not differ significantly on pretest mean response time on the Stroop task, F (2, 68) = 2.03; $p = 0.14$, showing the groups were not different before they started. The groups also did not differ significantly on an ANCOVA comparison of posttest reaction time by group with pretest as a covariate, F(2, 64) = 0.5, $p = 0.61$, $\eta2 = 0.02$, indicating no group improved more than any other. There is no evidence that playing 3 hours of *Lumosity* games focused on attention or flexibility skill improved performance on Stroop reaction time scores beyond a control group. The results for errors and response time on the Stroop task are consistent with the specific transfer theory (hypothesis 1) but not the other two hypotheses.

Discussion

None of the measures analyzed indicate that training for 3 hours with *Lumosity* games improves cognitive performance any more than an inactive control group. Although we do not have game performance data to assess whether participants at least improved within the games, all results from this study suggest any learning that occurs during game play does not transfer to performing the same cognitive skill in a non-game environment (hypothesis 2: specific transfer of general principles) or to performing a different skill in a non-game environment (hypothesis 3: general transfer).

The current study faced a number of challenges, most notably the training schedule. The duration of training was relatively short—3 hours of focused game playing—due to the limited amount of time our participants could give to our study. Future studies should aim to have participants train with the games for more cumulative hours. The training sessions were also somewhat long, in that a consumer would not normally spend a full hour playing the same 5 games repeatedly. Lumosity recommends consumers play for 15 minutes a day about 5 days a week. A study that could replicate this schedule would be able to more accurately assess the efficacy of this product.

Experiment 2

Experiment 2 is a randomized controlled experiment aimed at determining the cognitive consequences of playing *Lumosity* games for a long period (i.e., 15 to 20 hours spread over up to 80 sessions) with short play periods in each session (i.e., 15 minutes).

Method

Participants. The participants were recruited from the students and the surrounding community at a university in the western United States. Participants who successfully completed the training received a year-long personal subscription to the unrestricted Lumosity website as compensation. Overall 46 women and 45 men between the ages of 18 and 39 participated in the study (mean age = 21.2, SD = 4.12). Of the 91 people who started the study, 51 participants successfully completed the requirements of the study. Of these finished participants, 15 were assigned to play attention games using *Lumosity*, 16 were assigned to play flexibility games using *Lumosity*, and 20 were assigned to an inactive control group.

Materials. All cognitive tasks and all but one game are the same as those used in Experiment 1. We collaborated with the Human Cognition Project (HCP), the research branch of Lumosity, in order to accommodate the at-home training design. HCP provided customized test profiles for all active participants that restricted access to the five games assigned to their condition and provided information to the researchers about individual game performance and progress. HCP also provided the activation codes used as compensation for participants who completed their training. No financial incentives were provided to the researchers by HPC and no input was given beyond the planning phase over the course of this collaboration.

Cognitive tasks. The same cognitive battery from Experiment 1 was administered in Experiment 2; participants took the UFOV, WCST, Change Detection, and Stroop tasks before and after their training.

Flexibility games. The flexibility training group was asked to play the same 5 games used in Experiment 1: *Ebb & Flow, Disillusion (new), Color Match, Brain Shift*, and *Brain Shift Overdrive*.

Attention games. Four of the games from Experiment 1 were included in Experiment 2: *Eagle Eye, Observation Tower, Top Chimp*, and *Space Junk*. For full descriptions of these games see Study 1. One game was changed from Experiment 1 to Experiment 2. *Birdwatching* was replaced with *Star Search*, a newer attention game that is more dissimilar from the other games than *Birdwatching. Star Search* presents the player with a variety of somewhat similar shapes and one distinct shape. The player's goal is to identify the unique shape among the array. A screenshot of the game is shown in Figure 11.

Survey. Two surveys were given to the participants: one at the start of the study (pre-questionnaire) and one at the end (post-questionnaire). The pre-questionnaire asked demographic questions as well as questions about participants' video game history, such as how many hours a week they typically play action video games, casual video games, and brain training games. The post-questionnaire asked participants if they felt as though their attention skill, flexibility skill, or overall intelligence had improved on a Likert scale ranging from 1-5. In the active group version of the survey it also asked participants if they had enjoyed the games and whether they would continue to play them.

Procedure. Participants were randomly assigned to one of the three conditions (attention, flexibility, or inactive control) upon arriving at a psychology lab on a college campus. Participants were then given a consent form explaining what was expected of them and a pre-questionnaire that asked about their gaming habits and familiarity with *Lumosity*. After the consent form and pre-questionnaire were completed, the experimenter administered each of the four pretests in the following order: UFOV, WCST, Change

Detection, and Stroop. Once the tasks were complete the participants were given login information and instructions for their limited *Lumosity* profile for attention games (if they were in the attention group), flexibility games (if they were in the flexibility group) or no opportunity to play *Lumosity* games (if they were in the control group). This first lab visit lasted about 30 minutes.

The participants in the active training groups (attention and flexibility) were instructed to login to their limited *Lumosity* profiles 5 days a week for 15 minutes at a time (the length of one session) with the goal of finishing 20 hours of training in 16 weeks. To accommodate unpredictable schedules, participants were permitted to play up to 3 sessions in a sitting in order to stay on track with their goal of 5 sessions per week, but they were encouraged to stick to the prescribed schedule. Every week the participants who did not reach their goal of 5 total sessions per week received a reminder email from the researcher. The control participants were simply told to avoid *Lumosity* games and return to the lab after 16 weeks. The active participants completed an average of 73.5 sessions over an average of 18.5 weeks, resulting in an average of 18.3 hours of Lumosity game training. While this falls short of the 20 hour goal, game performance data shows that all participants had long since plateaued by the end of their training, so we feel confident that the amount of training was sufficient.

Once the active participants had reached at least 15 hours of training they were contacted and a follow-up session was scheduled. Control participants were emailed 15 weeks later to schedule their follow up session. Average time between pretest and posttest for participants was 130 days. In the final lab session, all participants were given the same battery of tests from the first session in the same order: UFOV, WCST, Change Detection, and Stroop. After the tests participants completed the post-questionnaire inquiring whether they felt as though they had improved since the last session. Finally, all participants who reached this point were given an activation code for unrestricted access to the full *Lumosity* website for the next year.

Results

Scoring. The main dependent measures for the UFOV, WCST, and Stroop remained the same as in Experiment 1. For the Change Detection Task accuracy was included in addition to reaction time as a dependent measure. Gain scores were calculated by subtracting the expected smaller value from the expected larger value. In the case of accuracy, pretest was subtracted from posttest. In the case of errors and reaction time, posttest was subtracted from pretest.

To establish participants' previous experience with games, their responses on a 5-point Likert scale to all survey questions about previous gaming experience were added together. These game scores were divided with a median split into a high game experience group and a low game experience group.

To establish participants' starting performance in the first game sessions they played, an average of their Lumosity Performance Index score (LPI, a standardized game score provided by Lumosity) for the first 3 sessions was computed. The same was done with the last three sessions that each participant completed. To establish their improvement, their first session score was subtracted from their final session score. A median split was then performed on this difference to create a high game improvement group and a low game improvement group.

To establish the high and low attention performance groups, the pretest scores for each of the four attention measures (UFOV accuracy, UFOV reaction time, Change Detection accuracy, and Change Detection reaction time) was standardized to reflect relative participant performance. We averaged each of these four z-scores together to create an overall attention z-score, and then performed a median split to create a high attention performance group and low attention performance group. The same was done for each of the four flexibility measures (WCST errors, WCST reaction time, Stroop errors, and Stroop reaction time). The z-score for each participant was averaged across the four measures, and a median split was performed on this standardized average to create a high flexibility performance group and low flexibility performance group.

Are the groups equivalent before the training? In order to assess whether participants improved with training, we must first establish that our groups were the same before training began. To do so we compared all eight pretest variables across the three treatment groups using analysis of variance (ANOVA). These tests revealed there were no significant preexisting attention differences among the groups on performance on UFOV accuracy, F (2, 84) = 2.76, $p = 0.07$; UFOV reaction time, F (2, 84) = 1.69, $p = 0.19$; Change Detection accuracy, F (2, 85) = 0.17, $p = 0.85$; and Change Detection reaction time, F (2, 85) = 0.31, $p = 0.74$. Likewise, there were no significant preexisting flexibility differences among the groups on performance on WCST errors F (2, 86) = 0.72, $p = 0.49$; WCST reaction time F (2, 86) = 0.07, $p = 0.93$; Stroop errors, F (2, 83) = 0.52, $p = 0.60$; or Stroop reaction time, F (2, 84) = 0.59, $p = 0.56$. There was no significant difference among the groups in mean age, F (2, 88) = 0.38, $p = 0.68$. Chi Square analyses showed that the three treatment groups did not differ significantly in the proportion of men and women, χ^2 (2) = 0.029, $p = 0.99$, and in the number of participants with high versus low previous gaming experience, χ^2 (2) = 1.46, $p = 0.48$. We conclude that there is not evidence that the groups differed in the basic characteristics of age, gender, gaming experience, and pretest performance by group.

Are there differences between participants who finished and participants who dropped out? Since a large number of participants failed to complete the study (n = 41), we wondered if there were any differences between the people who successfully completed the training and those who dropped out. To answer this question, we performed the same series of tests described above. While most results were not significant, we did find that those who dropped out (*M* = 592.28 *SD* =173.51) performed significantly faster than those who finished (*M* = 513.95, *SD* = 160.62) in pretests of UFOV reaction time, *F*(1, 85) = 4.763, *p* = 0.03. We also found that the mean age of participants who completed the study (*M* = 21.96, *SD* = 4.6) was significantly higher than of participants who dropped out (*M* = 20.2, *SD* = 3.2), *t*(89) = -2.06, *p* = 0.04, *d* = 0.44. We have no explanation for why those who dropped out should perform better on one of our attention measures. If it was a matter of not finding the tasks challenging enough, we would expect to see a higher dropout rate among participants with more gaming experience, which was not the case, χ^2 (1) = 0.01, *p* = 0.93. While we were surprised that such a small mean difference in age was significant, it is plausible that older participants felt a stronger sense of responsibility to the study, or that younger participants became more easily overwhelmed with the demands of school work over the course of the study.

Overall these results indicate that we can be confident that any differences we see between the groups in our posttests are due to the training intervention and not due to preexisting group differences. However, these results also indicate that we should be cautious to interpret the results as only applicable to the population of people who are willing to train consistently with *Lumosity*.

Do participants improve on the games? The major focus of this study is on the cognitive consequences of playing *Lumosity* games. If specific transfer theory is the most accurate model of transfer, we would expect trained participants to improve on the games, but to perform equal to untrained controls on the posttests. First we had to establish that *Lumosity* games do relate to the skills being measured. To establish this we divided participants into high and low-performing groups based on their pretest scores on the attention tasks (UFOV and Change Detection) and the flexibility tasks (WCST and Stroop) and analyzed the groups according to their average standardized score on their first three Lumosity game sessions. We found that participants who scored high on cognitive pretests achieved higher early game scores than participants who scored low on cognitive pretests, both for attention tasks, *F* (1, 29) = 4.59, *p* = 0.04, *d* = 0.95 (High: *M* = 1010.5, *SD* = 156.90; Low: *M* = 888.8, *SD* – 88.88), and for flexibility tasks, *F* (1, 29) = 6.57, *p* = 0.02, *d* = 0.85 (High: *M* = 1018.8, *SD* = 164.40; Low: *M* = 877.22, *SD* = 165.20). This confirms that the skills used in the games are related to the flexibility and attention skills being measured.

To establish whether participants who were assigned to play *Lumosity* games improved on the games we compared the average standardized score for the first three game sessions to the average standardized score for the last three game sessions with a paired samples t-test. There was a significant improvement between the first and last three sessions a participant played, *t*(30) = -18.99, *p* = 0.00, *d* = -3.40 (First: *M* = 959.46, *SD* = 165.23; Last: *M* = 1649.39, *SD* = 179.57). Not only did players improve on the games, the effect size was quite large (greater than 3 standard deviations). This indicates that the players learned very well in the context of the game, which is consistent with all three theories of transfer. Specific transfer theory would predict that improvements will be limited to the games, and learning will not transfer to our non-game measures of attention and flexibility. Thus, we must compare the groups' improvement from pretest to posttest on each measure to establish whether specific transfer, specific transfer of general skills, or general transfer theory best fits the results.

Do participants improve on the skill the games are training? Table 2a shows the mean scores (and standard deviations) on pretest, posttest, and gain for each group on the four attention measures, and Table 2b shows the mean scores (and standard deviations) on pretest, posttest, and gain for each group on the four flexibility measures. If the specific transfer of general skills theory is correct, we would expect (a) participants who played attention games to improve on measures of attention (UFOV and Change Detection) more than those whom played flexibility games or no games at all and (b) participants who played flexibility games to improve on measures of flexibility (WCST and Stroop). To test these predictions we ran ANCOVAs on posttest score with pretest score as a covariate and followed up with LSD post-doc tests (with $p < .05$) where appropriate. The method chosen to eliminate outliers was the extreme studentized deviate (ESD) or "Grubbs' test" (Grubbs, 1969). In order to perform ANCOVAs on data sets that are not normally distributed, some variables were converted to rank order before analysis (Connover & Iman, 1981). The variables that were analyzed by rank order due to skew were UFOV accuracy and WSCT reaction time.

First, concerning attention tests, there was no evidence that the attention group showed greater posttest scores or gains on attention tests. There was no significant difference among the groups on UFOV accuracy, Change Detection accuracy, or Change detection reaction time. An ANCOVA of posttest scores using pretest scores as a covariate revealed that there was a significant difference between groups on UFOV reaction time, $F(2, 44) = 3.43$, $p = 0.04$, but a Fisher's LSD post-hoc test indicated that the flexibility participants improved significantly more than the attention participants or the control participants, which did not differ significantly from each other. Thus, in four out of four attention measures, the attention group did not show greater posttest scores than the flexibility group or control group, yielding no support for the idea that skills trained in *Lumosity* transfer to non-game contexts (hypothesis 2).

Second, concerning flexibility tests, there was evidence that the flexibility group outperformed other groups on one of the four measures--Stroop reaction time. An ANCOVA on posttest score with pretest score as a covariate found a significant difference among the groups on reaction time on the Stroop task, F (2, 43) = 6.93, p = 0.00, $\eta_p^2 = 0.24$, and a post hoc LSD test revealed that the flexibility group had improved more than the attention or control groups. It should be noted that the Stroop test is very similar to one of the games (Color Match) played by the flexibility group. On three of the four flexibility measures (WCST accuracy, WCST reaction time, and Stroop accuracy), ANCOVAs on posttest score with pretest score as a covariate found no significant differences among the groups.

The same pattern of significant results was obtained when we ran ANCOVAs on gain scores with pretest score as a covariate. Overall, based on ANCOVAs, there is not strong evidence that training a skill with Lumosity games can improve that same skill in contexts outside of the game (i.e., this happened for one of the eight measures of cognitive skill), and thus, not strong support for the specific transfer of general skills theory.

Do participants improve in skills other than those trained by the games? If the general transfer theory is correct, we would expect participants who played attention games to improve on both the attention and flexibility measures relative to the control group, and we would expect the flexibility group to improve on both the attention and flexibility measures relative to the control group. Only one of the eight comparisons (described above and summarized in Table 2b) matched this pattern: the flexibility group shows more improvement than the other groups on UFOV reaction time, which is a measure of attention. This does not yield strong evidence for the idea that exercising one skill in game playing will result in improvements in other cognitive skills outside the game (hypothesis 3).

Supplementary analyses. Due to the wealth of survey and game data available to us, several other questions arose that begged answers.

Is training more effective for participants who improved most on the games? A potential explanation for the inconsistent improvement between groups could be that some people may be subject to a ceiling effect, i.e., they may already be performing at the top of their capacity. To explore this possibility, we hypothesized that if we compared the performance of those who improved most from the first to the last sessions of the games to the control participants, any potential training effects would be more apparent and the high game improvement group would show consistent improvements compared to the control group. A description of how we created the median split that isolated the high game improvement group can be found above in the scoring section. A t-test comparing the high game improvement group (M = 0.17, SD = 0.73) to the low game improvement group (M = -0.07, SD = 0.26) on the standardized attention gain scores revealed no significant difference between groups, t (13) = 0.94, p = 0.49. Those who had improved most on the games were not significantly more likely to improve on the attention skill overall. A second t-test comparing the high game improvement group (M = -0.32, SD = 0.49) to the low game improvement group (M = -0.20, SD = 0.57) on the standardized flexibility gain scores revealed no significant difference between groups, t (14) = 0.46, p = 0.65, high improvement, low improvement. Those who improved most on the games were not significantly more likely to improve on the flexibility skill overall.

Is subjective improvement associated with actual improvement? We hypothesized that participants who reported feeling as though they had improved more would have actually improved more. After the posttest we had participants indicate how much they felt they had improved in each domain on a 5-point Likert scale. A median split was created from this subjective rating to create a high subjective improvement group and a low subjective improvement group in both the attention and flexibility domains. An ANOVA comparing a high and low subjective attention improvement by standardized gain scores on the attention tasks revealed no significant difference between the high subjective attention improvement group (M = 0.08 SD = 0.60) and the low subjective attention improvement group (M = -0.10 SD = 0.47), F(1, 48) = 1.40, p = 0.24. Those who reported more improvement in the attention skill were no more likely to have actually improved on attention tasks, which contradicts our hypothesis. Likewise, an ANOVA comparing high and low subjective flexibility improvement by standardized gain scores on the flexibility tasks revealed no significant difference between the high subjective flexibility improvement group (M = -0.03 SD = 0.54) and the low subjective flexibility improvement group (M = 0.01 SD = 0.59), F (1, 48) = 0.06, p = 0.80. Those who reported more improvement in the flexibility skill were no more likely to have actually improved on flexibility tasks.

It remains possible that subjective improvement was an accurate assessment of game performance, if not an accurate assessment of skill improvement. To test this possibility, we correlated subjective improvement by game improvement, but found no correlation [r = -0.18, p = 0.33], indicating that subjective improvement was not based on an accurate judgment of game performance improvement.

Does subjective improvement differ by group? We hypothesized that the attention and flexibility participants would feel as though they had improved on all skills more than the control participants but that the attention and flexibility participants would not differ from each other in reported subjective improvement on any skill. Both analyses supported our hypothesis. An ANOVA with subjective attention improvement as the dependent measure and treatment condition as a between subjects factor indicated a significant difference among the groups, F (2, 47) = 14.5, $p = 0.00$, $\eta_p^2 = .38$ (attention: $M = 4.20$, $SD = 0.77$; flexibility: $M = 3.41$, $SD = 1.05$; control: $M = 2.42$, $SD = 1.02$). A Tukey post hoc test (with $p < .05$) revealed that the attention group and the flexibility group both reported significantly more attention improvement than the control group, but not more than each other. Likewise, comparing subjective flexibility improvement unveiled a significant difference by group, F (2, 47) = 7.90, $p = 0.00$, $\eta_p^2 = 0.25$ (attention: $M = 43.73$, $SD = 0.80$: flexibility: $M = 3.69$, $SD = 0.70$: control: $M = 2.63$, $SD = 1.20$. A Tukey post hoc test (with $p < .05$) revealed that the attention group and the flexibility group both reported significantly more flexibility improvement than the control group, but did not differ significantly from each other. These results indicate that participants who trained with *Lumosity* games felt as though they had improved both their attention skill and their flexibility skill more than participants who did not play games, regardless of which skill they trained, in spite of the fact that did not show strong skill improvement.

Discussion

The results of Experiment 2 do not provide strong evidence that the skills targeted in *Lumosity* games transfer to non-game contexts. In support of specific transfer theory, all active participants improved on the games, and on six of the eight measures the active groups did not improve more than the control group. This finding validates that spending time playing the game allows players to get better at playing the game.

In support of the specific transfer of general skills theory, the flexibility group improved more than the attention or control group on Stroop reaction time. Because of the strong similarities between the Stroop task and the Color Match game, the flexibility group's improvement can be interpreted as support for either the specific transfer of general skills theory or the specific transfer theory. However, seven of the eight measures of cognitive skill, playing *Lumosity* games for an extended period did not result in better improvements in the target cognitive skill than in the control groups.

In support of the general transfer theory, the attention group improved more than the flexibility or control group on UFOV reaction time. However, on the other seven measures of cognitive skill improvement, playing *Lumosity* games for an extended period did not result in better improvements in the non-targeted cognitive skill than in the control groups. We conclude that there is not strong support for the idea that playing *Lumosity* games targeting attention or flexibility skills is a highly effective way to improve performance on tasks requiring those skills outside the game.

General Discussion

Empirical Contributions

The first study, which trained participants for 3 hours spread over 4 sessions, did not yield positive evidence that *Lumosity* training causes improvements in cognitive skills. When the study was replicated using shorter training sessions spread over a much longer period of time – 18.5 hours over 18.3 weeks – the results also did not yield strong evidence that *Lumosity* training causes improvements in cognitive skills. Only two of the eight tasks showed domain-specific improvement for the trained groups, and one of these tasks is very similar to a Lumosity game with which some participants trained. We would expect to see consistent improvements in multiple measures if *Lumosity* training strongly improved cognitive skills outside of the game environment.

Methodological Contributions

The studies reported in this paper add to the body of literature on the cognitive consequences of existing commercial games to train cognitive skills. They both utilize random assignment, control groups, cognitive skill measures, and a healthy adult population (which is a rarity in cognitive training research). The incorporation of both an active and inactive control group further adds to the robustness of the design and can be used in the future to differentiate between training effects and placebo effects.

Theoretical Contributions

Three possible outcomes were hypothesized for these studies (1) specific transfer, which predicts that the attention group and flexibility group would improve on game performance; (2) specific transfer of general skills, which predicts that the attention group would improve on attention tests compared to the control group, and the flexibility group would improve on flexibility tests compared to the control group; and (3) general transfer, which posits that the attention group would improve on flexibility tests compared to control group and flexibility group would improve on attention tests compared to control group. Experiment 1 only provided evidence for specific transfer theory. Experiment 2 provided evidence for specific transfer theory and very modest evidence for the other theories (with no support in seven of eight tests for each theory). Overall, this study does not provide strong support for the idea that cognitive skills learned in a brain training game transfer to other cognitive skills or even to the same cognitive skill in non-game contexts.

Practical Contributions

The studies described in this report suggest that *Lumosity* brain training games may be effective in improving some cognitive skills within the game context, but these improvements are limited and may not transfer broadly beyond the game. Consumers should be wary of investing too much time and money if they expect tangible improvements, but if one enjoys the games there is no harm in devoting enough time to them to potentially see some limited return on investment. Tangible improvements aside, the present study suggests that regardless of if and how much a *Lumosity* user improves, they will feel as though they have improved. That subjective improvement in itself may have benefits future studies could explore.

Limitations

The current studies faced a number of challenges, most notably compliance and attrition. Of the 91 participants who started the long-term study, 40 did not complete it. The participants who did complete the study often failed to complete all 5 sessions in a given week, and had to be sent email reminders to make up for the missing game time. While the lack of strict compliance is a limitation, in a sense it does not weaken our findings. A normal consumer will very rarely be able to adhere to the 5-day-a-week regimen *Lumosity* prescribes due to the unexpected nature of life, as our participants found. Even with clear instructions and incentives, it was difficult for them to accommodate training with such regularity; what chance do normal consumers have of doing any better?

Future Directions

Do the effects last? Improving a skill is not all that useful if the improvement does not last very long. Periodic follow-ups with the completed participants could reveal if improvements last and if so for how long.

Is there a ceiling effect? A composite view of the literature suggests that brain training games may be more effective in cognitively impaired populations (Klingberg et.al., 2005, Kesler et. al. 2010, Finn & McDonald, 2011). This suggests that there may be a ceiling effect preventing cognitively healthy people from improving their performance any more. The plateau in game improvement we found also supports the notion that there is a peak level of performance. Alternatively, the plateau could be due to an inability for the games to get any more difficult, and given enough challenge players would continue to show improvement. This alternative is supported by the finding that those who improved most in the games were not more likely to improve on the tasks when isolated and compared to the control group. Future studies should explore if brain training games are more effective in lower-performing populations and if the capacity for improvement requires ever more challenging levels of play.

Is there a placebo effect? The survey results indicated that participants who played *Lumosity* games in the present study felt as though they had improved in all skills regardless of actual improvement or skill trained. To what extent is this subjective feeling of improvement related to expectation of improvement? Future studies should explore the placebo effect as a possible explanation for the subjective sense of improvement and the limited task performance improvement seen in the present study.

Can brain training games be improved? Knowing that cognitive improvement is possible with brain training games, the next step is maximizing their efficacy. Future studies should attempt a value-added approach (Mayer, 2014), taking existing brain training games and altering them to include elements that should theoretically enhance their efficacy, then comparing them to the original games. Potential added elements to explore include increasing the variety of the gaming environments for a given skill. Green and Bavelier

(2012) suggest too many trials over too few contexts could increase the specificity of the training. Perhaps decreasing the trials and increasing the variety of environments and game mechanics for a given skill will improve the likelihood of transfer.

The results point to the value of heeding the call for rigorous scientific research on the effectiveness of brain training games. Establishing how, when, and for whom games can improve cognitive skills can provide exciting possibilities for education and recreation, as well as valuable insight into the nature of learning and transfer.

Acknowledgement. The company that produces *Lumosity* provided the training profiles for the attention and flexibility groups, and provided a complimentary one-year subscription to participants who completed the study. The authors have no financial interest or connection with *Lumosity*.

References

Berg, E. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*, *39*, 15-22. doi: 10.1080/00221309.1948.9918159

Conover, W. J., & Iman, R. L. (1981). Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics. *The American Statistician, 35*(3), 124. doi:10.2307/2683975

Edwards, J. D., Vance, D. E., Wadley, V. G., Cissell, G. M., Roenker, D. L., & Ball, K. K. (2005). Reliability and validity of useful field of view test scores as administered by personal computer. *Journal of Clinical and Experimental Neuropsychology*, *27*, 529–543. doi:10.1080/13803390490515432

Finn, M., & McDonald, S. (2011). Computerised cognitive training for older persons with mild cognitive impairment: A pilot study using a randomised controlled trial design. *Brain Impairment*, *12*(3), 187–199. doi:10.1375/brim.12.3.187

Golden, C. J. (1978). *A Manual for the Clinical and Experimental Use of the Stroop Color and Word Test*. Chicago: Stoelting. Available at: http://nsuworks.nova.edu/cps_facbooks/47

Green, C., & Bavelier, D. (2012). Learning, attentional control, and action video games. *Current Biology*, *22*(6), R197–R206.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, *11*(1), 1. doi:10.2307/1266761

Hardy J. L., Drescher D., Sarkar K., Kellett G., & Scanlon M. (2011). Enhancing visual attention and working memory with a web-based cognitive training program. *Mensa Research Journal*, *42*(2), 13–20. doi: 10.1371/journal.pone.0134467

Hardy, J. L., Nelson, R. A., Thomason, M. E., Sternberg, D. A., Katovich, K., Farzin, & Scanlon, M. (2015, September 5). Enhancing cognitive abilities with comprehensive training: A large, online, randomized, active-controlled trial. *PLoS ONE, 10*(9).

Hulme, C., & Melby-Lervåg, M. (2012). Current evidence does not support the claims made for CogMed working memory training. *Journal of Applied Research in Memory and Cognition*, *1*(3), 197–200. doi:10.1016/j.jarmac.2012.06.006

Kane, M. J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working memory capacity and the top-down control of visual search: Exploring the boundaries of "executive attention". *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*(4), 749–777. doi:10.1037/0278-7393.32.4.749

Kesler, S., Hadi Hosseini, S. M., Heckler, C., Janelsins, M., Palesh, O., Mustian, K., & Morrow, G. (2013). Cognitive training for improving executive function in chemotherapy-treated breast cancer survivors. *Clinical Breast Cancer*, *13*(4), 299–306. doi:10.1016/j.clbc.2013.02.004

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., . . . Westerberg, H. (2005). Computerized Training of Working Memory in Children With ADHD-A Randomized, Controlled Trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(2), 177-186. doi:10.1097/00004583-200502000-00010

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

Melby-Lervåg, M., & Hulme, C. (2012). Is Working Memory Training Effective? A Meta-Analytic Review. *Developmental Psychology*, *49*(2), 270–291. doi:10.1037/a0028228

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods, 222*, 250–259. doi:10.1016/j.jneumeth.2013.10.024

O'Neil, H. F., & Perez, R. S. (2008). *Computer games and team and individual learning.* Amsterdam: Elsevier.

Owen, A. M., Hampshire, A., Grahn, J. a, Stenton, R., Dajani, S., Burns, A. S., Ballard, C. G. (2010). Putting brain training to the test. *Nature*, *465*(7299), 775–8. doi:10.1038/nature09042

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., … Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology. General*, *142*(2), 359–79. doi:10.1037/a0029082

Robbennolt, J. K. (2016, September). Brain games: Helpful tool or false promise? *Monitor on Psychology*, *47*(8), p. 18.

Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012). Working memory training remains a work in progress. *Journal of Applied Research in Memory and Cognition*, *1*(3), 217–219. doi:10.1016/j.jarmac.2012.07.009

Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. *Computers & Education*, *80*, 58–67.

Simons, D. J., Boot, W. R., Charbness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do brain training programs work? *Psychological Science in the Public Interest, 17*(3), 103-186.

Tobias, S., & Fletcher, J. D. (Eds.). (2011). *Computer games and instruction.* Charlotte, NC: Information Age Publishing.

Wouters, P., & van Oostendorp, H. (Eds.). (2017). *Instructional techniques to facilitate learning and motivation of serious games.* New York: Springer.

Appendix

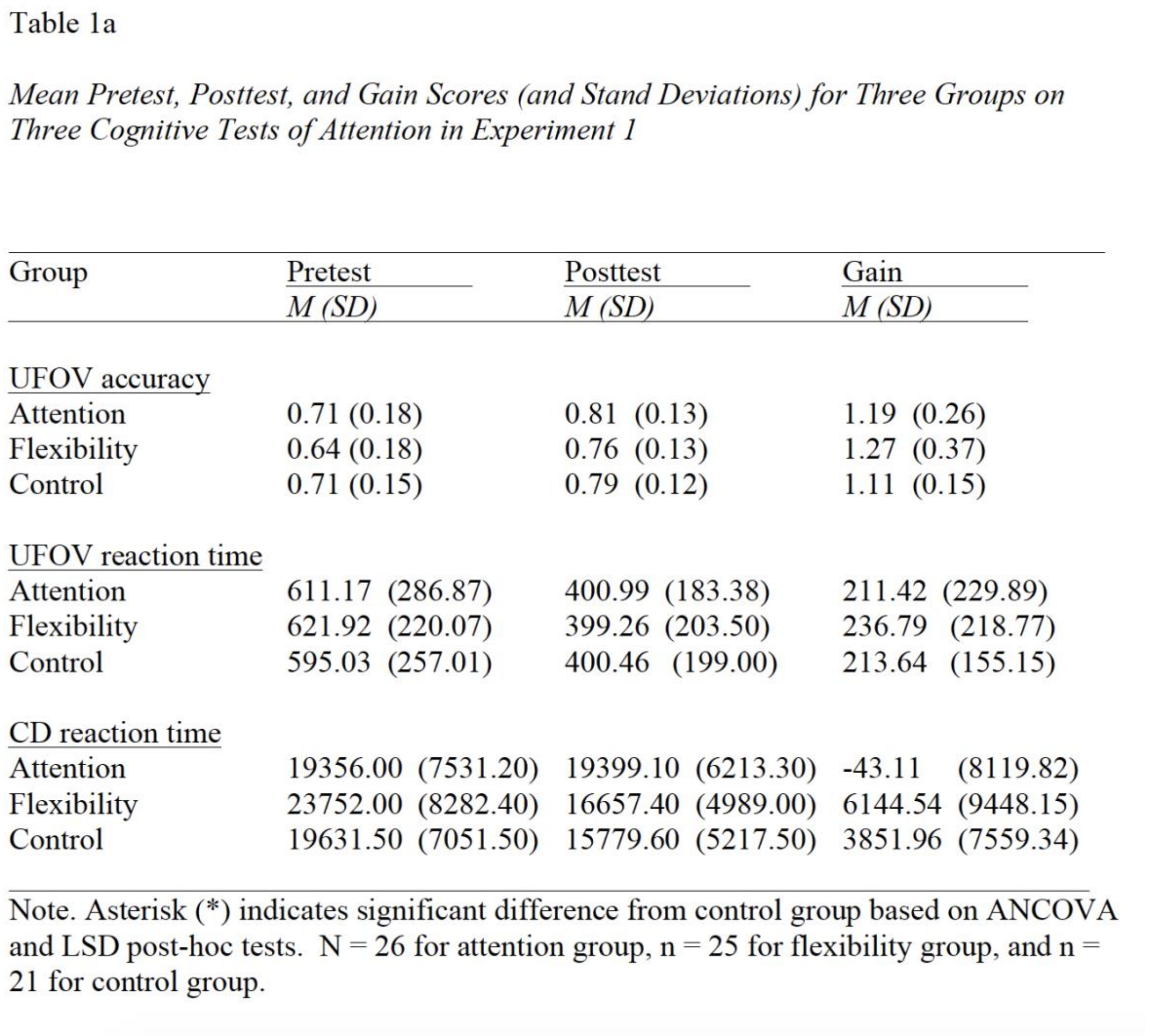


Table 1a

Mean Pretest, Posttest, and Gain Scores (and Stand Deviations) for Three Groups on Three Cognitive Tests of Attention in Experiment 1

Group	Pretest *M (SD)*	Posttest *M (SD)*	Gain *M (SD)*
UFOV accuracy			
Attention	0.71 (0.18)	0.81 (0.13)	1.19 (0.26)
Flexibility	0.64 (0.18)	0.76 (0.13)	1.27 (0.37)
Control	0.71 (0.15)	0.79 (0.12)	1.11 (0.15)
UFOV reaction time			
Attention	611.17 (286.87)	400.99 (183.38)	211.42 (229.89)
Flexibility	621.92 (220.07)	399.26 (203.50)	236.79 (218.77)
Control	595.03 (257.01)	400.46 (199.00)	213.64 (155.15)
CD reaction time			
Attention	19356.00 (7531.20)	19399.10 (6213.30)	-43.11 (8119.82)
Flexibility	23752.00 (8282.40)	16657.40 (4989.00)	6144.54 (9448.15)
Control	19631.50 (7051.50)	15779.60 (5217.50)	3851.96 (7559.34)

Note. Asterisk (*) indicates significant difference from control group based on ANCOVA and LSD post-hoc tests. N = 26 for attention group, n = 25 for flexibility group, and n = 21 for control group.

Table 1b

Mean Pretest, Posttest, and Gain Scores (and Stand Deviations) for Three Groups on Three Cognitive Tests of Flexibility in Experiment 1

Group	Pretest *M (SD)*	Posttest *M (SD)*	Gain *M (SD)*
WCST errors			
Attention	9.23 (4.35)	7.56 (4.94)	1.58 (5.70)
Flexibility	9.04 (4.25)	8.32 (4.17)	0.72 (3.55)
Control	8.57 (6.57)	7.29 (0.12)	1.29 (6.70)
WCST reaction time			
Attention	1844.77 (504.57)	1536.21 (429.78)	308.56 (466.55)
Flexibility	1851.45 (521.00)	1334.03 (283.79)	517.43 (463.55)
Control	1663.23 (461.07)	1368.39 (235.52)	294.84 (334.95)
Stroop errors			
Attention	14.08 (9.55)	11.46 (6.97)	2.35 (5.47)
Flexibility	17.13 (16.32)	14.24 (16.08)	2.20 (6.28)
Control	12.80 (7.39)	10.10 (3.89)	2.50 (6.89)
Stroop reaction time			
Attention	769.20 (112.10)	727.59 (82.78)	43.66 (76.07)
Flexibility	824.62 (251.83)	719.98 (138.07)	104.65 (259.65)
Control	736.82 (110.09)	686.45 (89.82)	48.59 (53.33)

Note. Asterisk (*) indicates significant difference from control group based on ANCOVA and LSD post-hoc tests. N = 26 for attention group, n = 25 for flexibility group, and n = 21 for control group.

Table 2A

Mean Pretest, Posttest, and Gain Scores (and Stand Deviations) for Three Groups on Four Cognitive Tests of Attention in Experiment 2

Group	Pretest *M (SD)*	Posttest *M (SD)*	Gain *M (SD)*
UFOV accuracy			
Attention	0.91 (0.12)	0.94 (0.04)	0.04 (0.08)
Flexibility	0.93 (0.04)	0.94 (0.03)	0.01 (0.04)
Control	0.93 (0.04)	0.93 (0.04)	0.00 (0.02)
UFOV reaction time*			
Attention	521.88 (173.09)	434.26 (156.39)	87.61 (101.43)
Flexibility	515.97 (170.19)	360.05 (83.18)	155.93 (128.63)
Control	493.76 (154.23)	430.01 (100.29)	62.76 (163.49)
CD accuracy			
Attention	0.94 (0.09)	0.94 (0.06)	0.00 (0.10)
Flexibility	0.93 (0.08)	0.94 (0.08)	0.01 (0.08)
Control	0.92 (0.08)	0.92 (0.08)	0.01 (0.07)
CD reaction time			
Attention	21459.49 (8113.25)	17824.58 (4886.76)	3634.91 (6179.73)
Flexibility	20357.00 (4780.30)	17455.93 (7004.91)	2901.08 (6825.95)
Control	20734.63 (5743.06)	17375.24 (5748.21)	3359.39 (8422.77)

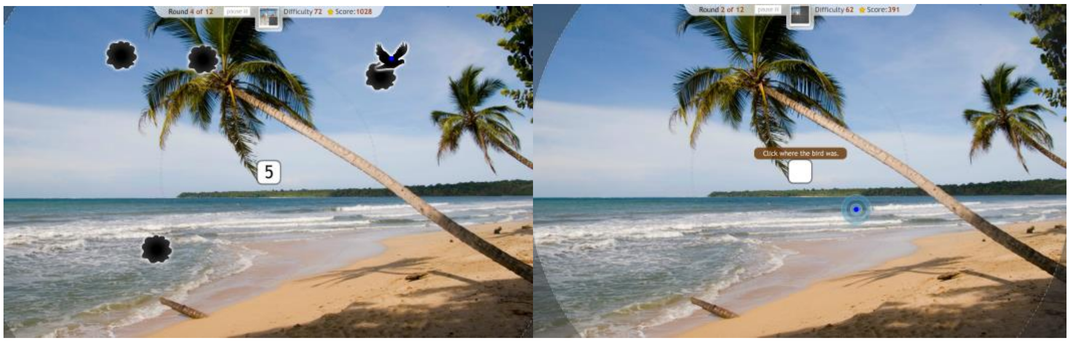
Note. Asterisk (*) indicates significant difference from control group based on ANCOVA and LSD post-hoc tests. N = 26 for attention group, n = 25 for flexibility group, and n = 21 for control group.

Table 2b

Mean Pretest, Posttest, and Gain Scores (and Stand Deviations) for Three Groups on Four Cognitive Tests of Flexibility in Experiment 2

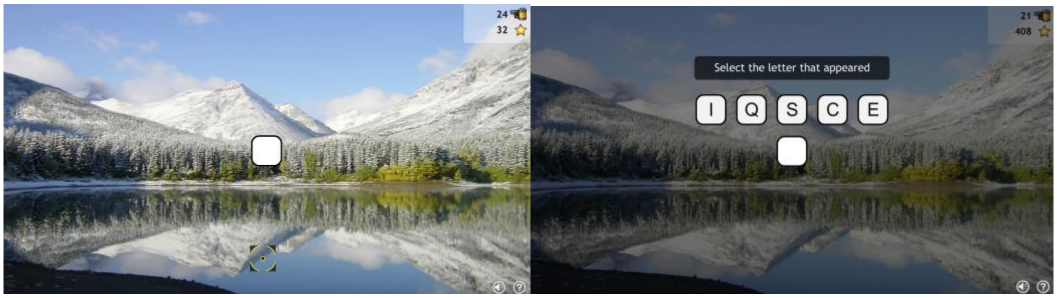
Group	Pretest *M (SD)*	Posttest *M (SD)*	Gain *M (SD)*
WCST errors			
Attention	8.13 (4.02)	6.80 (3.43)	1.33 (3.62)
Flexibility	6.75 (1.39)	5.81 (1.11)	0.94 (1.95)
Control	6.00 (1.87)	6.65 (2.09)	-0.65 (2.67)
WCST reaction time			
Attention	1737.69 (334.65	1413.99 (175.23)	323.70 (242.18)
Flexibility	1795.78 (697.50)	1407.62 (335.05)	388.16 (716.44)
Control	1949.22 (413.06)	1673.76 (399.69)	275.46 (397.25)
Stroop errors			
Attention	10.20 (3.67)	10.60 (5.62)	-0.40 (6.80)
Flexibility	11.93 (10.11)	8.79 (4.64)	3.14 (7.75)
Control	12.11 (8.07)	8.89 (4.91)	3.56 (7.63)
Stroop reaction time*			
Attention	741.44 (85.08)	739.72 (108.35)	1.72 (58.09)
Flexibility	806.14 (133.25)	699.94 (99.89)	106.20 (64.30)
Control	808.22 (125.20)	779.75 (135.12)	36.44 (85.06)

Note. Asterisk (*) indicates significant difference from control group based on ANCOVA and LSD post-hoc tests. N = 26 for attention group, n = 25 for flexibility group, and n = 21 for control group.



Click where the bird was.

Figure 1. Screenshots from the *Lumosity* game “Eagle Eye”. Text reads “Click where the bird was”



Select the letter that appeared

Figure 2. Screenshots from the Lumosity game “Birdwatching”. Text reads: “Select the letter that appeared”.

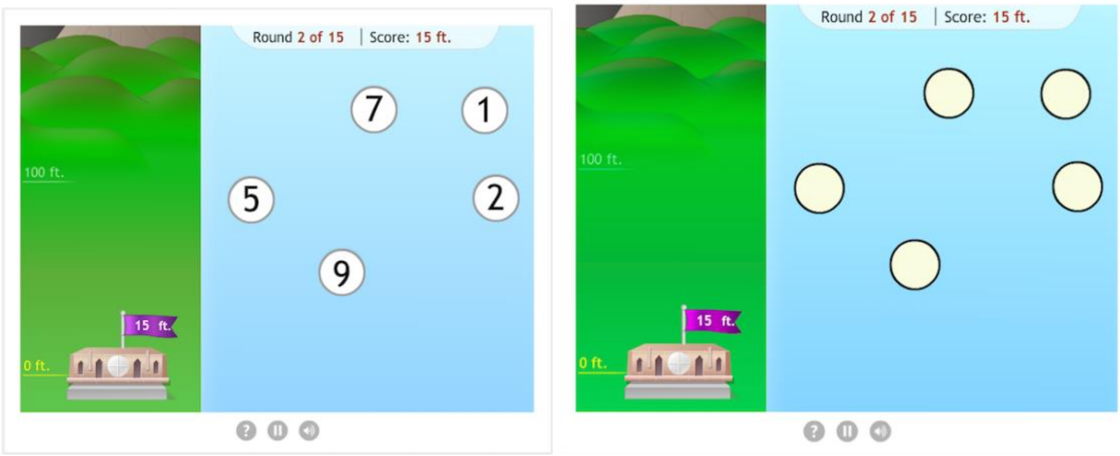


Figure 3. Screenshots from the Lumosity game “Observation Tower”

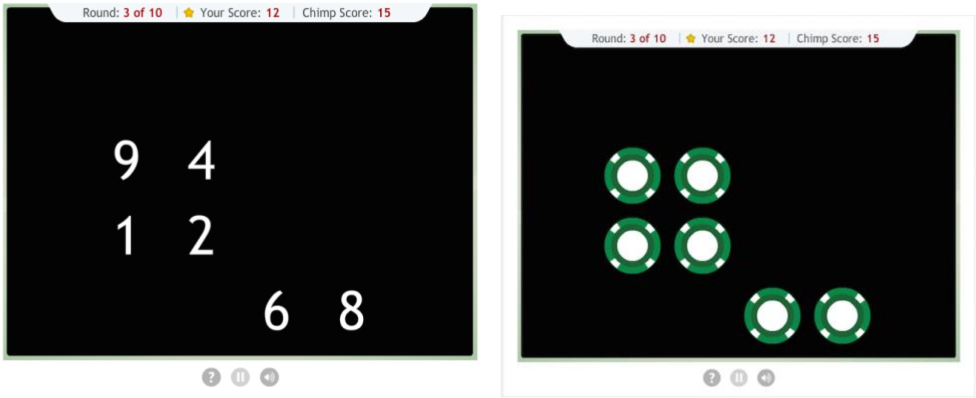


Figure 4. Screenshots from the Lumosity game “Top Chimp”

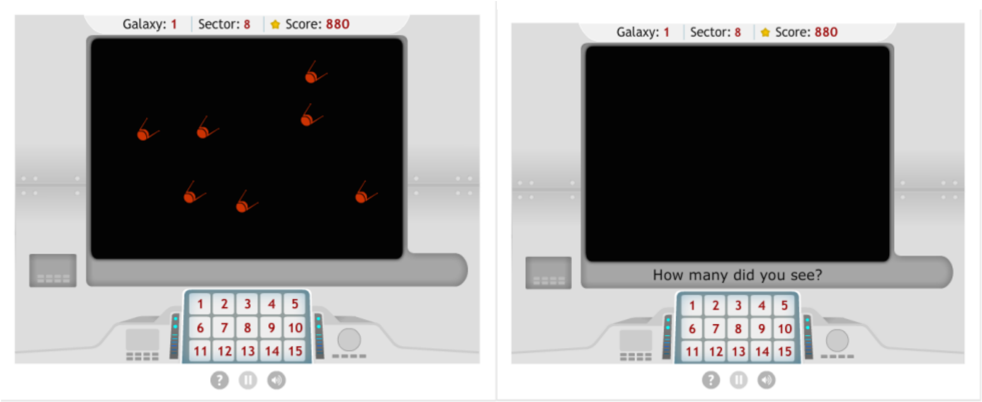


Figure 5. Screenshots from the game “Space Junk”. Text reads “How many did you see?”

Figure 6. Screenshots from the Lumosity game “Ebb & Flow”.

Figure 7. Screenshots from the Lumosity game “Disillusion: New”.

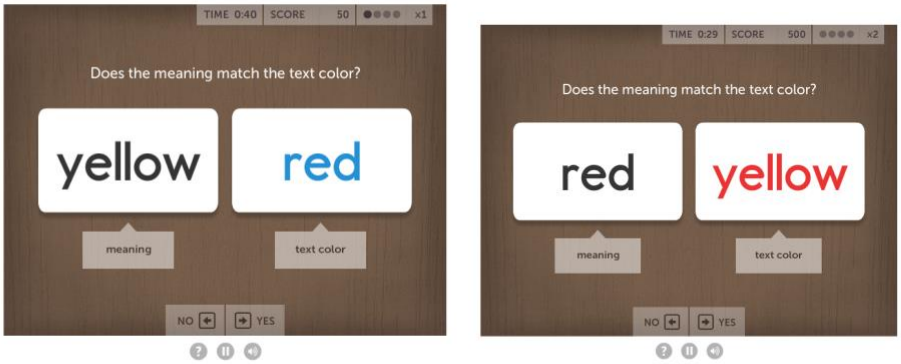


Figure 8. Screenshots from the game “Color Match”. Text reads: “Does the meaning match the text color?”

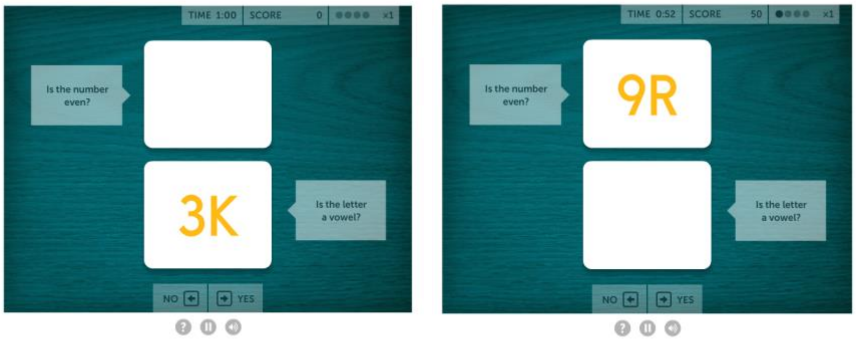


Figure 9. Screenshots from the game “Brain Shift”. Text reads: “Is the number even?/ Is the letter a vowel?”. After a few turns the cues go away.

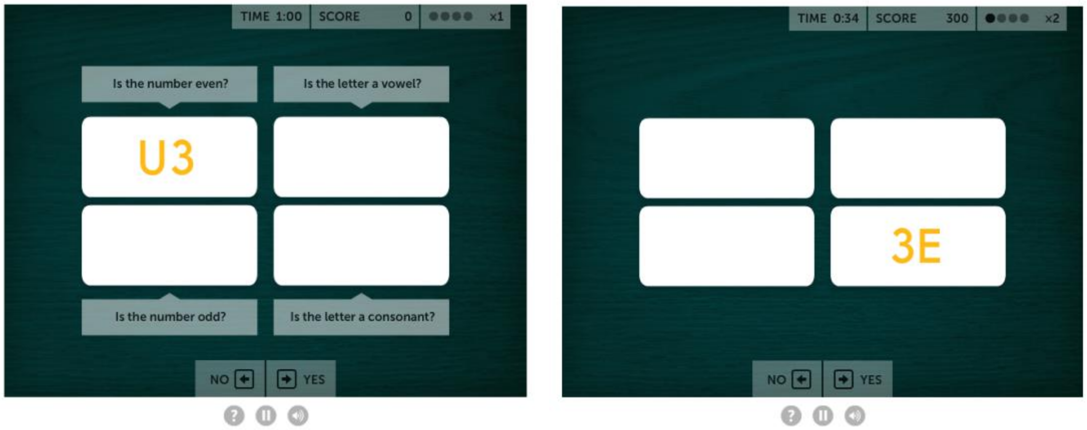


Figure 10. Screenshots from the game “Brain Shift Overdrive”. Text reads: “Is the number even?/ Is the letter a vowel?/ Is the number odd?/ Is the letter a consonant?”. Eventually the cues go away.

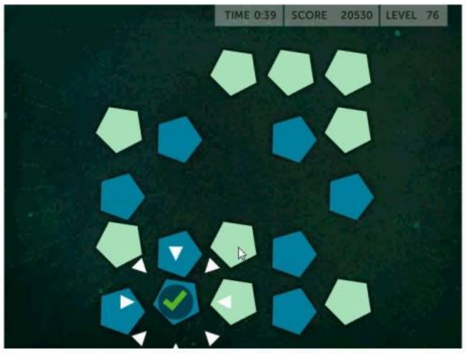


Figure 11. Screenshot from the game “Star Search”