Working memory and cognitive flexibility training reveals no relationship to fluid intelligence in college students

Sarah Luca¹,², Elliot Nauert¹, Keith Chichester¹, Jeannie Buckner³, Patrick Foo¹, Angeldeep W. Kaur⁴

¹Department of Psychology, University of North Carolina Asheville, Asheville, North Carolina, 28804; ²Department of Mathematics, University of North Carolina Asheville, Asheville, North Carolina, 28804; ³Department of Biology, University of North Carolina Asheville, Asheville, North Carolina, 28804; ⁴Department of Interdisciplinary Studies, University of North Carolina Asheville, Asheville, North Carolina, 28804

Recently, there has been an increased interest in cognitive training due to claims of widespread and transferable benefits of online brain training games. A growing body of literature supports the idea that working memory and cognitive flexibility are linked with fluid intelligence and academic success. The literature is less consistent on whether lasting improvements in cognition can be made through training these abilities. This study compared the effectiveness of cognitively challenging tasks, including Lumosity’s program, in building transferable abilities that contribute to improvements in fluid intelligence. To this end, cognitive performance by no-contact control participants was compared with that of two groups participating in either Flexibility-Focused Lumosity or Memory-Focused Lumosity trainings, and active control groups training in either Sudoku puzzles (alternate task control) or online trivia games (crystallized intelligence control). Measures of cognitive flexibility, memory and fluid intelligence were compared and showed significant improvements pre- and post-test, but not significantly greater improvement for any particular training group. These data suggest that the tested brain training programs are no more effective than any other cognitively engaging task in building transferable cognitive abilities.

Abbreviations:  ANOVA – Analysis of Variance; UNC – University of North Carolina; WAIS - Wechsler Adult Intelligence Scale

Keywords:  cognitive training; crystallized intelligence; fluid intelligence; working memory

Introduction

Interest in the development of learning, and how it can be maximized, has been on the rise. An increasing presence of personal electronic devices has been followed by a growing number of digital game-based products that claim to improve cognitive function and even increase overall intelligence. Game based training is appealing due to the ease of using games, their inherent motivating factors, and their accessibility on digital platforms. If game-based interventions are demonstrated to be effective in improving cognition, the findings would have implications in many domains including education, therapies for patients with neurological injuries, and combating age-related cognitive decline (Horowitz-Kraus, 2013; Lampit et al., 2015; Petrelli et al., 2015; van der Donk, 2015).

One of the front-runners of these programs is Lumosity, a commercially available product with 50 million subscribers (see e.g. Hardy et al., 2011), which claims to improve brain performance through the puzzles and games offered on its website and mobile application. Television advertising campaigns
Working Memory and Cognitive Flexibility Training Reveals No Relationship to Fluid Intelligence in College Students

2017

for Lumosity emphasize that their product is backed by neuroscience; however, most of the published scientific evidence that support these claims comes from Lumos Labs, the makers of the Lumosity gaming platform (Hardy et al., 2011; Sternberg et al., 2013). Studies testing the effectiveness of brain training platforms on older or cognitively impaired individuals have shown significant cognitive gains, but there is no clear consensus in the literature regarding the effectiveness of cognitive training in producing widespread, transferable gains in fluid intelligence in young, otherwise cognitively healthy individuals (Lampit et al., 2015; Petrelli et al., 2015; Jaeggi et al., 2008; Johnco et al., 2013).

Unlike crystallized intelligence, or knowledge of facts and figures which can be improved through training (Alloway and Alloway, 2009; Cunningham et al., 1975), fluid intelligence consists of a wide array of specific cognitive abilities that share complex interactions depending on current goals and context. Jaeggi et al. (2008) show evidence of improvements to measures of general fluid intelligence through training with computer-based working memory puzzles. However, this study did not include active control groups and used limited testing measures (Shipstead et al., 2010). Introducing an active control group (Harrison et al., 2013; Redick et al., 2013) and testing participants with an extensive battery of cognitive tests (Redick et al., 2013) provided no evidence of transfer to fluid intelligence. Thus, although Jaeggi et al. (2008) have suggested that fluid intelligence can be improved through training, other studies do not see the same levels of improvements (Au et al., 2015; Harrison et al., 2013; Jaeggi et al., 2008; 2014; Redick et al., 2013).

The present study was designed to test the effectiveness of brain training in young, otherwise cognitively healthy and engaged college students. This study focused on cognitive flexibility and working memory, two abilities that can be targeted for improvement in the Lumosity games. These abilities were chosen because they are central to success in the classroom, and there is some evidence of overlap between these cognitive abilities and general fluid intelligence.

Cognitive flexibility is the ability to switch back and forth between multiple goals, rules, or pieces of information. Current research supports the idea that cognitive flexibility is closely related to learning, at least in young school children (Bull and Scerif, 2001; Cartwright, 2002; Johnco et al., 2013; Kieffer et al., 2013; Yeniad et al., 2013). In studies looking at groups affected by various psychological disorders, cognitive flexibility has also been shown to be an important factor in emotional health (Brockmeyer et al., 2014; Wykes et al., 2003). Very few studies have looked explicitly at training cognitive flexibility, especially in the general population, but current research indicates that cognitive flexibility is a skill that can be enhanced (Brockmeyer et al., 2014; Glass and Maddox, 2013; Masley et al., 2009; Wykes et al., 2003). However, there is little evidence to suggest that the relevant training would result in widespread improvements in cognition, and in fact, cognitive flexibility may not be correlated with general intelligence at all (Friedman et al., 2006).

In the current study, the researchers seek to further clarify the effects of exercises that target particular cognitive abilities of general cognition. Specifically, several forms of digital game-based training strategies including the Lumosity training program are compared. This study focuses on the effects on working memory, cognitive flexibility, and fluid intelligence after exposing college students
(ages 18-24) to these brain training games. Many of these brain training platforms are advertised as providing training benefits for all age groups. If improvements are seen in this group in particular, whose members are assumed to be at the peak of their cognitive performance, an effect would likely be seen in all other age groups as claimed by the advertising.

Comparing an active training group to a no-contact control group may show effects resulting from the difference between cognitively engaged participants and cognitively stagnant participants, rather than effects resulting from the nature of the training itself (Au et al., 2015; Harrison et al., 2013; Jaeggi et al., 2008; Melby-Lervåg and Hulme, 2013; Redick, 2013; 2015). In this study, in addition to the inactive, no-contact control group, two distinct active control groups are used. Given the extensive body of literature supporting the idea that fluid intelligence is generally an untrainable cognitive skill, it is hypothesized that any improvements that may arise from training will not be significantly different from practice effects.

**Material and Methods**

**Participants**

This research was approved by UNC Asheville’s Institutional Review Board. All participants provided their written informed consent. Participants were recruited from UNC Asheville’s undergraduate student body. New groups were recruited at the beginning of each semester via flyers that advertised a “brain training” study, in-class sign-up sheets, and UNC Asheville’s online psychology research participation shell. Participants were offered “Psychology and Life” event participation credits, a requirement for most psychology classes at UNC Asheville, and were entered into a drawing to win a $20 gift certificate to a local ice cream parlor to promote adherence to the training requirements. All participants were given an informed consent form to sign before participation in this study and told they could withdraw at any time for any reason.

A total of 117 participants, aged 18-24 years, were recruited over the course of four semesters. Thirty six participants were excluded from the study either due to voluntary withdrawal or failure to complete a sufficient number of training sessions. The total number of participants who completed the experiment was 81, n = {15, 17, 17, 17, 15} for no-contact, alternate task, crystallized intelligence, memory, and flexibility groups, respectively.

**Training groups**

Participants were randomly placed into one of five different groups, including a no-contact control group that required participants to not complete any form of digital brain training exercise outside of their classwork for the six week testing period. The remaining four groups included two active control groups (crystallized intelligence “trivia” and alternate task “Sudoku”) and two experimental groups (Lumosity “Memory” and Lumosity “Flexibility”).

In the first active control group, participants completed online trivia games at the site, Sporcle.com. A group that is explicitly training crystallized intelligence should not show any improvements to fluid intelligence beyond those effects that may arise from general engagement and stimulation. Thus, any improvements to fluid intelligence in the other training groups must show improvements beyond those of the crystallized intelligence group.

Participants in the second active control group completed Sudoku puzzles regularly. Sudoku involves many elements of cognition and intelligence including spatial and numerical reasoning, making the puzzles an appropriate paradigm to which to compare Lumosity’s program (Grabbe, 2011; Lee et al., 2008).

The two experimental groups required participants to prioritize either memory or cognitive flexibility training as per Lumosity instructions. To mimic a real Lumosity training experience, participants were able to choose any of the games available within each category for their training using accounts that were set up by the researchers.

Regardless of which training group they were randomly assigned to, all participants were required to complete their assigned cognitive
exercises for 20 minutes a day for 3-5 days a week, as prescribed by Lumosity.com. The researchers kept in weekly contact with participants to maintain their prescribed exercises. The statistical program SPSS was used to analyze these data.

**Pre- and Post-Tests**

Participants were given a battery of tests to measure working memory, cognitive flexibility, and fluid intelligence before and after six weeks of training. These tests: memory span, Stroop, paper folding and matrix reasoning are detailed below.

**Memory span task**

Variants of the memory span task have consistently been used to measure changes in memory capacity resulting from cognitive training, especially those trainings focused specifically on working memory (Alloway, 2012; Bull and Scerif, 2001; Harrison et al., 2013; Jaeggi et al., 2008). Each trial consists of a sequence of digits, letters or words that are presented on the screen for one second each. At the end of a sequence, buttons labeled with digits, letters or words appear containing a label for each item presented in the sequence, including those not presented. The participant is asked to click the buttons in the same order the items appeared in the sequence. If the participant clicks the buttons of each item in the sequence in the correct order, the trial is marked correct. For correct trials, the next time a sequence is presented with those items the length of the list increases by one. An incorrect trial causes the length to decrease by one. In this task, each participant completed 30 trials, 10 for each type of sequence (digits, letters and words). The average number of correct responses for each type of sequence was used to calculate an overall average correct response score. The task was obtained from the program CogLab 2.0.

**Stroop task**

The Stroop Task is a commonly used task in psychological studies with more than 700 related articles relating to attention, conflict, decision making, and automaticity since J.R. Stroop’s original dissertation (MacLeod, 1991). In this study, the scores on the Stroop task are meant to index the level of cognitive flexibility in participants. The classic Stroop task and/or variations of it have been used to measure cognitive flexibility in many studies before, either independently or as a component of a battery of tasks (Glass et al., 2013; Johnco et al., 2013). Each participant begins a trial by pressing the spacebar and a fixation dot appears in the middle of the window. After less than a second of staring at the dot, a word for one of three different colors appears on the screen, RED, GREEN, or BLUE, with each of the font colors of these words presented in either red, green, or blue. The participant is asked to identify the font color of these words as quickly as possible. An incorrect response is presented again at a later trial. This task contained 48 trials, 24 with matching font color and word (congruent), and 24 in which they don’t match (incongruent). The difference in response times between congruent trials and incongruent trials were calculated and are presented as Stroop savings scores. Lower scores indicate a quicker, more accurate response, and thus better performance on the Stroop task. The Stroop task was also conducted using CogLab 2.0.

**Paper folding and matrix reasoning**

Fluid intelligence was measured with two tasks: the matrix reasoning subtest of the Wechsler Adult Intelligence Scale-III, and the paper folding task. Both tasks have been successfully used together to measure general or fluid intelligence (Johnson et al., 2004; 2008). Matrix reasoning tasks have been widely used in measuring general intelligence (Raven, 2003). In this task, participants are shown a set of patterns arranged in a grid, with one cell of the grid empty. They must choose from five possible choices the shape that would best complete the pattern if it were placed in the empty grid. Paper folding is used to assess an individual’s spatial reasoning abilities which have been linked to general cognitive factors (Boonen et al., 2014; Ekstrom et al., 1976; Tosto et al., 2014). In the paper folding task, participants are shown an image of a piece of paper that has been folded in a specific pattern, and then had a hole punched through it. They then choose from a selection of
images what the paper would look like if it were unfolded again, based on how the holes would be arranged. In both tasks, scores were calculated by summing the number of correct responses.

**Results**

One-way analyses of variance (ANOVA)s indicated that there were no significant difference in scores among groups at pre-testing (paper folding: $F(4, 76)= 1.263, p > 0.05$; matrix reasoning: $F(4, 76)= 0.482, p > 0.05$; memory span: $F(4, 76)= 0.684, p > 0.05$; Stroop: $F(4, 76)= 1.279, p > 0.05$). Scores in each of the four tasks were analyzed by 5x2 mixed-effect ANOVAs, with a between subjects factor of training group (no-contact, crystallized intelligence, alternate task, memory, or cognitive flexibility), and a within subjects factor of testing time (pre- or post-training). Results from the ANOVAs are compiled in Table 1.

**Table 1. Significance testing results for cognitive measures.**

<table>
<thead>
<tr>
<th>Group</th>
<th>Testing Time</th>
<th>Group x Testing Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F$</td>
<td>$p$</td>
</tr>
<tr>
<td>Paper Folding</td>
<td>0.572</td>
<td>0.684</td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>0.342</td>
<td>0.849</td>
</tr>
<tr>
<td>Memory Span</td>
<td>0.534</td>
<td>0.711</td>
</tr>
<tr>
<td>Stroop</td>
<td>2.137</td>
<td>0.084</td>
</tr>
</tbody>
</table>

*Significant $p$-values

The ANOVAs revealed no significant interactions between training group and testing time, all $p > 0.05$. A main effect of Testing Time (all $p < 0.01$) was observed indicating improvements in performance from pre- to post tests for all cognitive ability tests except the Stroop test (as mentioned before, an improvement on the Stroop task is observed with a lower Stroop savings score due to quicker response times and better accuracy, see Table 1). However, there was no one group that showed significant improvement in the post-tests than any other. This is evidenced by the overlapping of gains scores, representing the amount of improvement observed from pre- to post-test (Figure 2).
Working Memory and Cognitive Flexibility Training Reveals No Relationship to Fluid Intelligence in College Students

2017

Figure 1: Average test scores for each training group. A main effect of Testing Time before (lighter color) and after training (darker color) show significant improvements ($p < 0.01$) for all cognitive ability tests except the Stroop test. Conversely, no main effect of training group for any task was observed, $p > 0.05$. a) Mean number correct responses in paper folding task. b) Mean maximum memory span length achieved in memory span task. c) Mean number of correct responses in matrix reasoning task. d) Mean difference in response time between congruent and incongruent trials in the Stroop task. Values represent mean $\pm$SEM. n=15-17.

Figure 2: Participant improvement between pre- and post-tests. Normalized mean gain scores for all groups in each task show that no one task is better at improving overall fluid intelligence. Raw scores were normalized to account for differences in scaling among the testing methods. Normalized scores for each group were calculated by the formula $(X - X_{min})(X_{max} - X_{min})$, where $X$ is each score raw score, $X_{min}$ is the minimum score for the group, and $X_{max}$ is the maximum score for the group. Values represent mean $\pm$SEM. n=15-17.
Discussion

This study revealed no significant improvements in working memory, cognitive flexibility, or fluid intelligence as a result of cognitive training programs (Figures 1, 2). While there were improvements in some measures of cognitive ability during the training period, none of the tested training programs appear to be more effective than any other. The gain scores in the experimental groups and the active control groups did not significantly differ from the gain scores of the no-contact control group. Additionally, no training-related interactions were observed for the memory and flexibility groups, since neither group experienced significant improvements in the memory span task and Stroop task respectively. This may indicate that the improvements observed were merely practice effects in the pre- and post-test tasks themselves (Figure 2). Another possibility is that participants experienced some level of cognitive enhancement that inherently occurs for college students over the semester as a result of their curriculum.

Numerous studies report training-related improvements to measures of general intelligence (Au et al., 2015; Jaeggi et al., 2008; Melby-Lervåg and Hulme, 2013; Redick, 2015; Redick et al., 2013). However, other researchers have reported no improvements when they have increased sample sizes, included active control groups, expanded the pre- and post-test measures, and used more rigorous statistical analysis (Harrison et al., 2013; Redick et al., 2013; Redick, 2015). Ideally, many different pre- and post-tests would be conducted to provide greater power, sensitivity, reduced error variance and reliability to measures of changes in intelligence (Redick, 2015). In order to create tests of manageable length, only subsets of full intelligence testing batteries were used in this study. In an earlier paper by Redick et al. (2013), 17 different tasks were used in the pre- and post-test batteries yet they did not detect significant training-related improvements to cognitive ability.

Inclusion of appropriate active controls is particularly important. Without them, any improvements observed in such studies could be attributed to a placebo effect, where participants would expect to do better after their training sessions and would in fact show improvements. These improvements might not be because of greater fluid intelligence, but could instead be attributable to other factors, such as greater vigilance during post-tests. In addition, using active controls allows us to determine whether a significant improvement is observed simply from doing a “cognitive task” rather than a “non-cognitive task”. Redick et al. (2013) found that both their active control and experimental group improved with practice on their respective tasks, yet there was no transfer to any of the cognitive abilities measured by their extensive battery of tests. In this study, two active control groups were included. One of the active control groups was specifically designed to train crystallized intelligence so that no improvements in fluid intelligence were expected to be observed. However, current results indicate that the crystallized intelligence group improved the most on the paper folding task, though this improvement was not statistically significant (Figures 1, 2). These results further emphasize the need for an active control group, as they suggest that improvements observed in previous studies are most likely due to a placebo effect. It is shown here that game-based cognitive training programs aimed at improving memory and cognitive flexibility do not produce significant transferable improvements to fluid intelligence in participants between the ages of 18 and 24 years.

Conclusion

The scientific community and regulatory entities seem to be coming to a common consensus on the true nature of commercial cognitive training programs. In 2014, an open letter from the Stanford Center on Longevity challenged companies that claim to provide effective brain training programs. Signed by a long list of esteemed scientists, the letter objects to the claims of brain training companies that their products offer scientifically backed solutions for combating cognitive decline (Stanford Center on Longevity and Max
Planck Institute for Human Development, 2014). Additionally, the Federal Trade Commission (FTC, 2016) recently charged Lumosity with deceiving its users by claiming that its products would improve cognitive health and subjected the company to a heavy fine. The data reported in this study provide experimental support for the perspective increasingly expressed by experts across the field; there is little evidence to suggest that cognitive training programs are uniquely effective in improving or protecting general cognitive ability.

Acknowledgements

The authors contributed equally to this work. The authors would like to thank the following alumni from the University of North Carolina at Asheville: Alex Schaeffer, Quentin Reynolds, Melissa Allen, and Abbey Allen. These individuals provided assistance with recruiting participants and collecting data. “The Hop” supplied the gift certificate. This research was supported by a grant from the University of North Carolina at Asheville Undergraduate Research Program.

References

Grabbe JW (2011) Sudoku and working memory performance for older adults, Act