TRANSFER EFFECTS IN ONLINE WORKING MEMORY TRAINING

A MASTER THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF SOCIAL SCIENCES AND HUMANITIES OF

KOÇ UNIVERSITY

BY

MUSTAFA KAYA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR

THE DEGREE OF MASTER OF ARTS

IN

PSYCHOLOGY

JULY 2014

ABSTRACT

Thus far, working memory training studies were always conducted in confined laboratory environments. This study aimed to carry out a working memory training procedure in a naturalistic online environment. After extensive piloting, online tests and a training task were developed using recent technologies such as HTML5 and KineticJS with cross-browser compatibility. Participants were randomly assigned either to the experimental group that received an adaptive training schedule or the control group that received a non-adaptive training schedule. To test the effects of the online training task and to measure near and far transfer, we developed the online versions of eight tasks forward digit span, backward digit span ,letter span, alphabet span, Stroop inhibition task, block design, Raven's Standard Progressive Matrices (RSPM), and verbal arithmetic task. Participants took a pretests and a posttest in two sessions. The suggested training regimen was 15 minutes in each session, four sessions a week during 6 weeks (46 days in total). Our initial sample consisted of 2,002 adults aged between 18 and 58 (M = 28.28). In the training task, a number of images were shown on a 3 × 3 grid and participants were asked to choose the correct images among nine images and drag them into the correct initial position. In the group with rotation, the grid rotated 90 degrees to the left or right 50% of the time, which required a visual transformation of information.

Our analyses showed that memory tasks, the arithmetic task, and the block design task had satisfactory test-retest reliabilities. Moreover, all the tests in this study showed appropriate clustering in the principal component analysis. The first component consisted of RTs in memory tasks, arithmetic and block design task. The second component had only accuracy measures in memory tasks. The third component had high component loadings from the Stroop task. The fourth component received the highest loadings from the RSPM and arithmetic accuracies. As for the training, in both groups we observed stabilization in maximum level scores after the fifth session. The training task had moderate correlations with backward digit span accuracy that measures storage and manipulation in working memory and arithmetic RT, which is a measure of fastness of response in verbal problem solving. Manipulation of information rather than passive storage was a better predictor of problem solving skills. At the end of the study, there was no difference between the experimental and control groups while we detected a significant difference between the contact (10 or more training sessions) and no-contact groups (less than 10 training sessions) in digit span differences and a marginally significant difference in letter span differences. Both of them were in the expected and desired direction. Although we had a low compliance level in our training, the training task has also a promising future as a measure of fluid intelligence, manipulation of information, and problem solving. We also studied the nature of attrition and compliance in our data including predictors of attrition. We provide future directions and suggestions about this fertile but untouched area of online testing and training of working memory.

Keywords: Online working memory training, working memory testing, adaptive training, online cognitive testing, online cognitive training, practical applications in cognitive psychology.

ÖZET

Şimdiye kadar, çalışma belleği eğitimi çalışmaları her zaman sınırlı laboratuvar ortamında yapılmıştır. Bu çalışmada çevrimiçi ve doğal bir ortamda çalışma belleği eğitimi prosedürü yürütülmesi amaçlanmıştır. Kapsamlı bir pilotlamadan sonra, çevrimiçi testler ve bir eğitim görevi tarayıcılar arası uyumluluk ile HTML5 ve KineticJS gibi son teknolojiler kullanılarak geliştirilmiştir. Katılımcılar, rasgele adaptif eğitim programı alan deney grubuna ya da adaptif olmayan eğitim alan kontrol grubuna atanmıştır. Çevrimiçi eğitim programını ve yakın ve uzak transfer etkilerini test etmek için, ileri ve geri sayı menzilleri, harf ve alfabe menzili, Stroop, blok tasarımı, Raven Standart Progresif Matrisleri (RSPM) ve sözel aritmetik testi olmak üzere sekiz çevrimiçi test hazırlanmıştır. Katılımcılar iki oturumda ön test ve art testleri almıştır. Önerilen eğitim rejimi her oturumda 15 dakika, 6 hafta boyunca dört oturum (toplam 46 gün) olmuştur. Başlangıç örneklemi yaşları 18 ve 58 arasında (*Ort.* = 28.28) 2002 katılımcıdan oluşmuştur.

Analizler, bellek görevleri, aritmetik görevi ve blok tasarımı görevinin tatmin edici test-tekrar test güvenirliğine sahip olduğunu göstermiştir. Ayrıca, bu çalışmadaki tüm testler ana bileşen analizinde uygun kümelenme göstermiştir. İlk bileşen bellek görevleri, aritmetik ve blok tasarımındaki görev tepki sürelerinden oluşmuştur. İkinci bileşende yalnızca bellek görevlerindeki doğruluk ölçümleri bulunmuştur. Üçüncü bileşen en yüksek bileşen yüklerini Stroop görevinden ve dördüncü bileşen ise RSPM ve aritmetik doğruluklarından almıştır. Eğitim görevinde ise, her iki grupta beşinci oturumdan sonra oturum maksimum düzeyi skorlarında bir istikrar gözlenmiştir. Eğitim görevi, depolama ve çalışma belleğinde işleme ölçen geri sayı menzili ve sözel problem çözmede yanıt hızının bir ölçüsü olan aritmetik görevi tepki süreleriyle orta düzeyde korelasyon göstermiştir. Problem çözme becerilerinin en iyi yordayıcısı ise edilgen depolama değil bilgi işleme kapasitesi olmuştur. Çalışmanın sonunda, deney ve kontrol grupları arasında herhangi bir fark tespit edilemezken, devam eden grupta (10 oturum ve üstü katılım gösteren) devam etmeyen gruba göre (10 oturumdan az katılım gösteren) sayı menzili farklarında anlamlı bir fark, harf menzili farklarında ise anlamlılığa yakın bir fark bulunmuştur. Her ikisi de beklenen ve istenen yönde olmuştur. Eğitimde düşük uyum düzeyi olmasına rağmen, eğitim görevi akışkan zekâ, bilgi işleme ve problem çözme ölçeği olarak da gelecek vaat etmektedir. Bunların yanı sıra, katılımcı kaybı ve uyumu da katılımcı kaybının yordayıcıları dahil olmak üzere incelenmiştir. Verimli ancak az çalışmış bir alan olan çalışma belleğinin çevrimiçi ölçüm ve eğitimi konusunda gelecek çalışmalar için doğrultular ve tavsiyeler sunulmuştur.

Anahtar Sözcükler: Çevrimiçi çalışma belleği eğitimi, çalışma belleği test etme, adaptif eğitim, çevrimiçi bilişsel test etme, çevrimiçi bilişsel eğitim, bilişsel psikolojide pratik uygulamalar.

DEDICATION

To my mother, my family, all the people I aim to help, and

music that made me go and kept me alive,

ACKNOWLEDGEMENTS

I would love to provide sincere thanks to all the participants, who showed interest in the project and provided their contribution,

My thesis advisor Aylin C. Küntay for her strong belief in me, understanding and encouragement, my readers İlke Öztekin and Zehra Çataltepe for their invaluable contributions, and my professors Esra Mungan for showing her enthusiasm and support, Sami Gülgöz for relying on me without doubts and providing whatever he could do to assist me, Kemal Kuşçu for his constant tracking and help, and Tarcan Kumkale for advising better methods,

My mother, my aunts, and my family for their appreciation of my work,

My supportive close friends Kaan Türkyılmaz for his endless patience and compassion, Zeynep Funda Yazıcı for her limitless availability, morale and corrections to my drafts, Yasemin Görür, Ayça Alaylı, Kerem Dülger, Beyza Bilal, Handan Titiz for making me realize the conditions and keeping me focused, Oğuzhan Güngör, Nazlı Altınok for her initializing ideas, Ceren Yalın, Erhan Özer, Serdar Metin, Cenan Tuna, Süen Çankaya, for encouraging me to program video games and Ezgi Aytürk,

My geeky, knowledgeable, and very assistive friends Çağatay Tuncer, Taha Bilge, Yiğit Kocagöz, Gülçin Akturan, Berke Gürkan, and Onur Küçük for their support in demographic ratings,

Peak Games for providing their sponsorship by providing an iPad for our participants and online announcements that helped us to boost our participation rate to a great extent,

İmge Özdemir and Özge Özdemir for their incredible hospitality during my initial data collection for piloting in Ankara, My lab co-workers Deniz Özkan Göktürk and Beyza Ateş Şen,

Yüce Okulları (Yüce Schools) that provided me the opportunity to work with the pupils, and school psychologist Fatma Başak Ulusoy for her comprehension of difficulties in data collection and providing support for it,

Metin Kara, for receiving a training about our program, applying it with their clients with autism and providing unprecedented feedback for our applications,

My psychotherapist Beti Hayim for caring and supporting me in lots of ways,

TÜBİTAK for financing my studies, Koç University for providing a safe territory to concentrate, Suna Kıraç Library personnel for being always positive, swift and assistive,

The New Boston (Buckly Roberts), PHP Academy, W3Schools.com, stackoverflow.com for creating marvelous and free online resources for enhancing my computer skills.

TABLE OF CONTENTS

ABSTRACT ... ii

ÖZET ... iv

DEDICATION ... vi

ACKNOWLEDGEMENTS ... vii

List of Tables ... xii

List of Figures ... xiii

INTRODUCTION ... 1

1.1. The Nature of Working Memory ... 1

1.2. Working Memory Training ... 5

1.2.1. Strategy Training and Domain General Training Programs ... 6

1.2.2. Adaptive Working Memory Training Paradigm ... 7

1.3. Challenges and Limitations in Previous Working Memory Training Studies ... 10

1.4. The Current Study ... 12

METHOD ... 14

2.1. Participants and procedure ... 14

2.1.1.1. Piloting Process ... 16

2.1.1.2. Demographic information ... 17

2.1.2. Procedure ... 19

2.2. Pre- and posttest measures ... 20

2.3. Training ... 29

2.3.1. Training Design ... 29

2.3.2. Positive feedback 32

RESULTS 35

3.1. Data Preparation..35

3.1.1. Data Entry and Preparation for Cognitive Tests 35

3.1.2. Data Preparation for Training Scores and Formation of Contact and No-Contact Groups 36

3.2. Analysis of Test Scores..36

3.2.1. Reliability of Test Scores..36

3.2.2. Correlations among Test Scores 39

3.2.3. Principal Component Analysis for Test Scores 43

3.3. The Nature of the Data and Understanding Attrition..46

3.3.1. Group Assignment and Attrition..46

3.3.2. Predicting Attrition from Pretest Scores 47

3.4. Training Scores 47

3.4.1. Correlations among Training Scores 47

3.4.2. Training Scores over Time..48

3.5. Correlations among Test Scores and Training Scores 51

6. Group Comparisons 52

DISCUSSION..55

4.1. Reliability of the Test Scores..55

4.2. Correlations among Test Scores 56

4.3. Training Scores 58

4.4. Effects of Training on Test Scores..59

4.5. Limitations and Future Directions 60

REFERENCES 63

APPENDICES 71

Appendix A. Demographic Questionnaire...71

Appendix B. Forward Digit Span Numbers...74

Appendix C. Backward Digit Span Numbers..76

Appendix D. Letter Span Letters 78

Appendix E. Alphabet Span Letters..80

Appendix F. Questions in Arithmetic Test 82

Appendix G. Images Used in the Block Design Task 84

Appendix H. Turkish and English Names of Animals Used in the Training Task..90

List of Tables

Table 1. Percentages of frequency of video game playing 17

Table 2. Percentages of ratings of experienced gamers in two dimensions 18

Table 3. Test-retest reliability of task scores 36

Table 4. Test-retest reliability of task scores in the no-contact group 37

Table 5. Pearson correlations among test scores in pretests 40

Table 6. Spearman correlations among test scores in pretests 42

Table 7. Component loadings for pretest variables 44

Table 8. Correlations of measures of training performance 47

Table 9. Correlations among test scores and training scores for participants who attended at least ten sessions 51

Table 10. Sample sizes, means and standard deviations for differences of digit spans in pre- and posttests 53

Table 11. Sample sizes, means and standard deviations for differences of letter spans in pre- and posttests 53

List of Figures

Figure 1. The flow of participants during the study 14

Figure 2. Participants were shown a plus sign before the beginning of each trial in span measures to ensure that all participants start from the same area on the screen 20

Figure 3. Span stimuli was shown in the middle of the screen 21

Figure 4. Digit span responses were received with a special screen keypad 21

Figure 5. Letter span and alphabet span tasks had a special screen keyboard composed of all letters used in the tasks 22

Figure 6. Four areas of block design task: (1) the top ribbon, (2) the source board (left), (3) the target board (middle), and (4) the help frame (right) 23

Figure 7. A sample screen from Raven's Standard Progressive Matrices; the top and large pattern has a missing part. There are six or eight answer options at the bottom of the screen 25

Figure 8. A screenshot from the arithmetic task. The question and keypad are centered on the screen and there is a timing bar in the right side of the screen 26

Figure 9. A sample screenshot from an incongruent trial in Stroop task. Four buttons at the bottom are the names of the colors as choice options. The text is "yellow" in Turkish but it is typed in green 27

Figure 10. A screenshot from the training task while waiting for the participant to respond (the top ribbon is open only when user input is allowed) 29

Figure 11. Screenshot of a feedback phase in the training session 30

Figure 12. A sample screenshot from the "boards" page displaying the badges, attendance information and shared achievements of other users. Names are hidden with rectangles for privacy of the participants 32

Figure 13. The scree plot showing eigenvalues for each component number 44

Figure 14. Maximum level in training session in each session for 6 participants in the experimental group 49

Figure 15. Maximum level in training session in each session for 17 participants in the control group 50

CHAPTER 1

INTRODUCTION

Working memory is defined as a memory system designed to handle storage and processing of task-relevant information that needs to be immediately available to cognitive operations (Shah & Miyake, 1999). Many researchers now believe that working memory capacity is closely related to other cognitive abilities such as reasoning, problem solving planning and inhibition. It is also the best single predictor of fluid intelligence (Colom, Flores-Mendoza, & Rebollo, 2003; Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004; Smith & Kosslyn, 2009). Since it is a core component in cognitive tasks, some researchers proposed that working memory has some plasticity and can be trained (Klingberg, 2010).

Brain training, especially working memory training, has received a great interest from researchers in the last ten years and has been extensively marketed to a wide range of populations (Holmes et al., 2010; Holmes, Gathercole, & Dunning, 2009; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Klingberg, Forssberg, & Westerberg, 2002; Mezzacappa & Buckner, 2010; Olesen, Westerberg, & Klingberg, 2004; Redick et al., 2012; Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2009; Westerberg et al., 2007). Although there are many working memory studies, to our knowledge, none of them has attempted to conduct this training outside the laboratory environment.

1.1. The Nature of Working Memory

The Atkinson-Shiffrin model, which is an early model of human short term memory, proposed three components of human memory (Atkinson & Shiffrin, 1968). The model suggested that information is first stored in sensory memory using sense organs for a very brief period of time, then, transferred to short-term memory and eventually long-term memory. A classic example of short-term memory is keeping a phone number to dial in our memory. This memory is limited both in capacity and time. There is a limit to the number of digits that can be held in the memory at once and the information will be lost unless it is transferred to the long-term memory or special techniques like rehearsal are used.

Later research suggested that these memory systems are not as linear as initially thought. To account for accumulated empirical results, Baddeley and Hitch (1974) developed a "three-component model of working memory". The concept of working memory differed from the short-term memory mainly by the fact that short-term memory refers to simple storage of knowledge; however, working memory refers to both storage and manipulation of information for a brief period of time. For instance, when a participant is asked to remember a number of words, the major active process is information storage. In contrast, when the participant is asked to put the words in alphabetical order and repeat them to the experimenter, the stored information needs also to be manipulated.

As its name suggested, the model is composed of three main components. The first one is an attentional controller component, which is called "central executive". This master component had two slave components mainly for the storage of information. One of these systems is allocated for speech-based information and it is called the phonological loop. The other one is for visual and spatial information and it is called the visuospatial sketchpad. There are two-way (encoding and retrieval) connections between these slave systems and the central executive but the slave systems does not have a direct connection. They are assumed to be capable of processing their own type of information or the other type of information but only if it is passed through central executive. In a later revision, Baddeley (2000) introduced a new set of components. He proposed that the first three components are fluid; that is, not very open to change and development. He added three crystallized components, which can be changed through learning. The first two of these crystallized systems are kinds of semantic information related to two slave systems with two-way connections to fluid components. They are called visual semantics for visuospatial sketchpad and language for phonological information. These two slave and crystallized systems are connected to each through episodic long-term memory. The revised model explains the interaction of previous learning and long-term memory with the fluid slave systems.

Baddeley's model strongly emphasizes the non-unitary nature of working memory because it suggests separable subsystems that interact with each other. In stark contrast, Cowan's Embedded-process Model (1999) proposes a unitary view with three subsystems of memory that are embedded in each other. The memory in the focus of the attention is embedded within a memory that is out of the focus (within the active part of long term memory) and in turn this subsystem is embedded within the inactive memory with retrieval cues (i.e., long term memory). Cowan's model is not domain-specific (cf. visuospatial sketchpad that processes *only* visual information and phonological loop that processes *only* phonologic information). Despite major differences from Baddeley's model, Baddeley (2012) held that this model is not incompatible with his model, the only difference being terminology and emphasis.

Another prominent account, Engle's Controlled Attention Model, working memory capacity is explained by controlled attention which is trainable and domain-free combined with potential domain-specific components, although there is no consensus among researchers neither about the number nor the types of these components (Engle, Kane, & Tuholski, 1999; Engle, 2002, 2010; Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2002). This account holds that working memory capacity is not related to the passive short-term storage ability, but it is "the capacity for controlled, sustained attention in the face of interference and distraction"; hence, controlled attention. Engle, Kane and Tuholski (1999) compared this controlled attention component to the central executive component in Baddeley's model. This account also emphasized the strong relationship between the controlled attention that provides a base for information manipulation with the fluid intelligence capacity. Studies of proactive interference (Kane & Engle, 2000), structural equation modeling of tasks of short-term memory, working memory and general fluid intelligence (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002), and contrasting low- and high-span participants in Stroop task (Kane & Engle, 2003) and dichotic listening (Conway, Cowan, & Bunting, 2001) provided the support for the accounts emphasizing the executive attentional nature of working memory capacity.

Apart from these accounts, early research studying the relationship between the linguistics skills such as reading comprehension and working memory capacity led researchers to propose accounts about the nature of passive short-term storage and information processing (i.e., manipulation of information in memory). There are three theories that explain the relationship between storage and processing. Processing Overlap theory suggested that the positive correlation between comprehension skills and reading span task scores is a result of the fact that processing component of the span task requires processes close to the ones required by comprehension tasks (MacDonald & Christiansen, 2002). Within this framework, Waters and Caplan (1996) held that processing and storage are entirely separate from each other. Daneman and Carpenter (Just & Carpenter, 1992) developed another theory, Resource Sharing theory, which maintained that processing and storage relies on a shared resource. Hence, there is a trade-off between storage and processing components. When one component is strained, performance in the other component is lower. The third is Time-Based Forgetting theory. This theory is based on findings concerning forgetting in working memory. It holds that the amount of processing does not only delimit/restrict the amount of information to be potentially stored but in addition, affects the retention of already stored information via decay or interference (for a brief comparison of these accounts, see Friedman & Miyake, 2004).

Considering the recent evidence in literature cited above showed that working memory capacity that predicts the performance in problem solving, general fluid intelligence and inhibition abilities was not related to storage capacity but the executive components of working memory, which can be called central executive, controlled attention or cognitive control depending on the theoretical approach, researchers speculated that increasing this information manipulation capacity may improve cognitive functioning a variety of tasks. In the next section, theoretical approaches to working memory training and previous studies with working memory training attempts will be discussed.

1.2. Working Memory Training

The ultimate aim of providing cognitive training is to improve the cognitive functioning of trainees in everyday contexts. Current researchers commonly follow a sequence of steps to test the effectiveness of a training protocol. To demonstrate benefits, first, participants should exhibit an increase in performance in the trained tasks. Second, they should perform better (e.g. greater span, faster reaction time) in tasks that are different than the trained ones but measure the trained abilities or closely related abilities. This step is called "near transfer". Third, they should show increased performance in tasks that require trained abilities as well as untrained processes. This kind of transfer is a more secure and stringent measure of training effects and this step is called "far transfer". For example, measures of fluid intelligence such as Raven's Progressive Matrices and some non-verbal subtests of intelligence tests, attentional control measures like the Stroop task are widely used as far transfer measures in working memory training studies.

1.2.1. Strategy Training and Domain General Training Programs

Approaches to working memory training can be grouped into two categories: (1) strategy training and (2) core training (Morrison & Chein, 2011). Strategy training, also called explicit training (Klingberg, 2010), targets only encoding and retrieval processes in short term memory. It focuses on specific strategies used by encoding or retrieval components of memory, for example, articulatory rehearsal and elaborative encoding as mnemonic techniques. For example, Turley-Ames (2003) instructed participants to repeat words to be remembered out loud until they were tested. The study showed that previously low-span participants benefit from rehearsal strategy. Although in some contexts such training may be useful in everyday life, Ericsson and Chase (1982) pointed out a severe limitation in this kind of training. One exceptional participant was able to remember up to 79 digits; however, he showed typical performance with non-numeric stimuli. Strategy training was shown to be material-specific and *was able to produce only near transfer, but not far transfer.* These strategies could not be used in tasks that are dissimilar to the trained tasks.

As opposed to strategy training, core training methodologies have a broader scope of effect. They target domain general processes; i.e., they aim to produce far transfer. Most of the current working memory training studies focus on core training and many of them are commercialized. Cogmed, JungleMemory and Lumosity are typical examples of this type of working memory training programs. A characteristic feature of these programs is employing many types of tasks instead of a single and specific training task. Although the list of tasks that are included in the Cogmed program are not fully disclosed to the public, studies that report using this common training program mention tasks that tap into separate types of working memory. For instance, Klingberg, Forssberg, and Westerberg (2002) used a visuospatial working memory task, a backward digit-span task, a letter-span task and a choice reaction time task. Interestingly, some tasks that are not specifically considered to be working memory tasks but are rather inhibition tasks such as choice reaction time task and go/no go tasks are commonly included in these studies and working memory training programs (e.g., Bergman Nutley et al., 2011; Olesen, Westerberg, & Klingberg, 2004; Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2009).

1.2.2. Adaptive Working Memory Training Paradigm

The term "adaptive working memory training" was recently introduced into the literature and many researchers (Alloway & Parker, 2011; Klingberg, 2010; Shipstead, Redick, & Engle, 2010, 2012) emphasized that training programs must be designed "adaptively". Given that participants start at different initial points or progress at different rates during a training protocol, training programs could be designed to flexibly adjust to the previous levels demonstrated by specific trainees. Adaptiveness of a task refers to adjusting the difficulty of the task according to recent scores received from the trainee. The idea behind this adaptation is that since working memory is assumed to be trainable, like a muscle, a task should challenge the participant at the boundary level. When the participant performs well at a level, the difficulty is increased and otherwise decreased. Hence, they continue to practice at the level that they are experiencing a difficulty in (Shipstead et al., 2010, p.249). However, in a later work, Shipstead, Redick, and Engle (2012) pointed to three factors that are critical to defining a working memory program as "adaptive". First, it should not be providing strategy (explicit) training but implicit and domain general training. Second, it should include exclusively working memory tasks. Third, the adaptation criterion should be participant's ongoing scores.

Researchers used extremely varied training and outcome tasks in core working memory training along with a large range of age and special groups. However, usage of inconsistent tasks resulted in incomparability of many studies and inconsistent findings. Even studies which employed Cogmed training tasks provided different sets and versions of tasks in different regimes (for the review, see Morrison & Chein, 2011).

Olesen, Westerberg, and Klinberg (2004) conducted two adaptive working memory training experiments that took five weeks each with young adults. In their first experiment, they trained participants with visuospatial working memory, backward digit span, and letter span tasks. They showed that training improved performance in Span board task, which is another measure of working memory. In the second experiment, they failed to show the same effect using Span board task as well as a digit span task but they detected differences in brain activity and attributed this to training. Authors suggested that the difference in behavioral data may be due to differences in procedures in two experiments. Klingberg, Forssberg and Westerberg (2002) used a similar training battery but also had a choice reaction time task in their second experiment with four young adults and 25 ADHD children and a similar training schedule. They showed training effects in both groups in Span board scores. Buschkuehl and Jaeggi (2010) trained old adults with spatial working memory and reaction time tasks. They showed training transfer in block span scores but not in digit span, verbal and visual free recall scores. Near transfer was shown in all tasks in other studies which used Cogmed tasks. Holmes and colleagues showed the effect with children with ADHD (Holmes et al., 2010) and working memory deficits (2009). Similar results were also obtained in a study with preschool children (Thorell et al., 2009) and with children with attention deficits (Mezzacappa & Buckner, 2010).

Far transfer from working memory training to other abilities includes general fluid intelligence, attentional control or inhibition abilities, and also arithmetic abilities. General fluid intelligence is strongly related to working memory capacity (Colom et al., 2003). In training studies, researchers obtained conflicting results in fluid intelligence. Studies that employed Raven’s matrices as a measure of fluid intelligence did not yield consistent results.

Many studies detected increases in Raven's Matrices scores with training (Klingberg et al., 2005, 2002; Olesen et al., 2004) as well as in other measures of fluid intelligence such as Block Design, which is a subtest of Wechsler Intelligence Scale for Children (Bergman Nutley et al., 2011). Westerberg and colleagues (2007) showed training related increases in Raven's matrices in young adults but not in patients with stroke. Thorell and colleagues also failed to show a working memory training effect in Block Design in preschool children. The inconsistencies may be resulting from different training regimes and components as well as the large variability of sample groups.

Since executive functions such as working memory and effortful control are both empirically and theoretically closely related (Zhou, Chen, & Main, 2011), researchers studied the effects of working memory training on attentional control and inhibitory abilities. A very common measure of inhibitory capacity is Stroop task. The conflicting results as in the case of fluid intelligence also emerged from research. Some studies detected a training effect in Stroop performance (Klingberg et al., 2005, 2002; Olesen et al., 2004) while other failed to show this (Wykes, Reeder, Corner, Wuliams, & Everitt, 1999). Similarly, in a study with preschool children, Thorell and colleagues (2009) failed to show changes in inhibitory abilities related to visuospatial working memory training.

Mathematics, like many complex cognitive abilities, requires many processes at once which include executive functions such as attention, working memory and planning. Research on the relationship between working memory and mathematics showed that they both rely on similar processes and working memory capacity is definitely and somehow related to mathematical skills (Raghubar, Barnes, & Hecht, 2010). However, the nature of this relationship is not well understood. A salient reason for this situation is that both are complex processes that rely on many other processes and functions. Also, there are many external factors that complicate the relationship such as age, skill level, the mode of presentation of the stimulus and the type of mathematical ability. Swanson and Beebe-Frankenberger (2004) showed that working memory still predicted math tasks even when controlling for speed, inhibition and general knowledge. They argued that regulation and/or manipulation of information is crucial to mathematical processes. To our knowledge, there is only one working memory training study which explored the effects of training on numerical skills. Holmes and colleagues (2009) trained children with working memory deficits in Cogmed tasks and they failed to show any training effect on Wechsler Objective Numerical Dimensions task.

Although there is a large number of studies that provide support for the success of adaptive working memory training (for a list of core training studies see Morrison & Chein, 2011, p. 51-52), researchers indicated some concerns about the validity of this support (e.g. Shipstead et al., 2012).

1.3. Challenges and Limitations in Previous Working Memory Training Studies

Previous literature pointed to several major caveats in working memory training studies. The first one is the extensive usage of passive or no-contact control groups (Morrison & Chein, 2011; Shipstead et al., 2010). To avoid methodological artifacts such as the Hawthorne effect, placebo effect, Pygmalion effect, and demand characteristics; the golden standard for training research is employing double-blind designs. However, most studies, especially the ones that were published in the early 2000s, had no-contact control groups (Morrison & Chein, 2011). For blind control groups in training studies, researchers were advised to employ active control groups (Redick et al., 2012; Shipstead et al., 2010, 2012). Using an adaptive working memory training paradigm, as Morrison and Chein (2011, p. 57) and Shipstead and colleagues (2012) advised, there are several studies which provided a non-adaptive version of the same task to the control group; that is, researchers excluded only the adaptiveness component of the training to create an active control group to keep the standards of accepting an evidence for effectiveness at a more conservative level (Holmes et al., 2009; Klingberg et al., 2005, 2002).

The second common criticism is that researchers used similar tasks (e.g. span tasks) both in the training and the post-training effects measurement (Morrison & Chein, 2011; Shipstead et al., 2010). Shipstead and colleagues (2012) argued that the increase in scores may be due to practice with span tasks rather than a genuine amelioration of working memory capacity. Hence, to overcome this limitation, Morrison and Chein (2011) suggested using a larger number of tasks including those unrelated to the training task to measure training and transfer effects.

Bryck and Fisher (2012) noted that the studies are mostly conducted in the laboratories. They argued that laboratory-based training lacks ecological or external validity and that there is not sufficient evidence to support that benefits of training extend to daily cognitive functioning or real world situations. Richmond, Morrison, Chein, and Olson (2011) conducted a computerized adaptive working memory training study with older adults and showed a difference in everyday functioning in one subtest of Test of Everyday Attention with only well-trained participants. Wykes, Reeder, Corner, Wuliams, and Everitt (1999) provided working memory training to adults with schizophrenia. They used two tasks to measure planning ability of the participants. They showed an increase in the training group in the modified six elements test, but not in the Tower of London. They also demonstrated an increase in self-esteem scores. Nevertheless, all studies to our knowledge used a strict training regime. Participants were asked to complete fixed-timed sessions in fixed intervals. To keep the ecological validity of the study, allowing participants some extra training for a limited time or number of trials may be important in addition to fixed training. Moreover, this extra-training may provide some variability in dosage. Dobrzenski (2009) proposed that duration and frequency of training may be manipulated to estimate the optimum schedule of training. Jaeggi, Buschkuehl, Jonides, and Perrig (2008) argued that dosage positively affects the amount of transfer; nonetheless, Redick and colleagues (2012) criticized their results due to their methodological and statistical choices and their failure to show that dosage is related to the amount of transfer.

1.4. The Current Study

Apart from carrying the working memory out of the laboratory, to real life conditions, the major goal of the current study was to address particularly the potential caveat of using no-contact groups. The current study was conducted with a training group and an active control group. These groups are divided into two smaller groups. One of the groups received training that focuses on only storage while the second group received training in both storage and processing. This was planned to allow some variance to the study of the relationship between the dosage of training and individual differences. There was no specific hypothesis about this relationship because previous literature did not provide sufficient information about these complex relationships. Therefore, this part of the study was more exploratory. However, in line with findings presented above, we formulated six hypotheses to be tested.

Our first hypothesis was that training effects would be observed in forward digit span, backward digit span, letter span, alphabet span tasks and Stroop task; that is, experimental groups would show a greater increase compared to the control groups. The second hypothesis was that increases in the scores of the problem solving tasks (Block Design and Raven's Standard Progressive Matrices) that are highly related to fluid intelligence would be greater in groups that practice storage and processing compared to the groups that practice only storage, considering that recent research suggested that passive storage is not related to problem solving and fluid intelligence but manipulation of information (Engle, 2002). Also control groups are not expected to show changes in problem solving tasks, because of not receiving a fully adaptive task at their boundary level. The third hypothesis was that increases in arithmetic scores would be greater in the storage and processing training group compared to its control. The other two groups were not expected to differ from each other. The fourth hypothesis was that increases in arithmetic scores would be greater in the storage and processing training group compared to its control. The other two groups are not expected differ from each other. The fifth hypothesis was that participants who had lower initial task scores would benefit more from training compared to participants with higher initial task scores because previous research suggested that participants with lower scores had greater increases because of their larger room available for improvement (Klingberg, 2010). The sixth hypothesis was that participants who scored higher in activity control, inhibition control and attention would benefit more from the training because they may be more devoted and alert to new information during training sessions and be less affected by environmental distractors.

CHAPTER 2

METHOD

2.1. Participants and procedure

2.1.1. Participants

We recruited 2002 participants online in a time period of approximately 13 months. Figure 1 shows the flow of participants during the study.

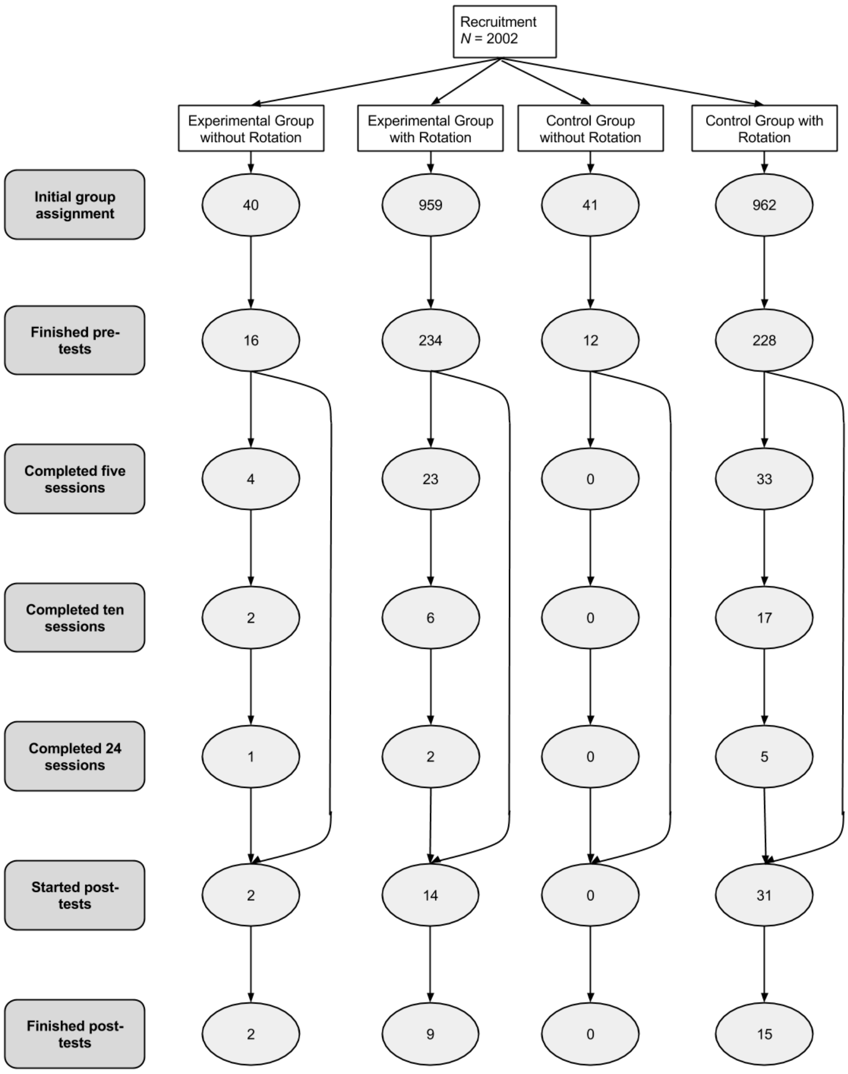


Figure 1. *The flow of participants during the study.*

The study was announced by news coverage in a popular science magazine (Ekici, 2014), several online game and news sites (for an incomplete list, see http://www.grittygames.org/public/media), social media posts, and email announcements of Koç University. Peak Games, one of the largest social gaming companies, sponsored the project. The company shared social media posts in peak hours of users and majority of participants volunteered through these announcements, especially from the quiz games called "Who Knows" and "Know and Conquer". Although we also received participants from card and board games like 'Batak' or 'Okey', the number was low. Because of this, we directed the announcements more towards quiz games. We also announced that one participant who complied with the suggested training schedule would win an iPad Mini. The device was provided as a result of sponsorship of Peak Games and was awarded to one of the participants with a raffle as announced before. Moreover, all participants who took the tests were awarded with their test scores after the study was completely terminated along with information about how to interpret these scores.

Participants were randomly assigned to groups when they signed up. Participants were informed about the procedure of the study before they completed the registration form. In the registration form, they selected a username and a password. Logins were managed by this username and password information. All passwords were hashed before writing them into the database to keep this information completely private and secure. We also collected their names, surnames and email addresses for future contact and appropriate addressing. Their emails were confirmed with a hyperlink sent to their addresses to ensure the integrity of data collection and future contact.

After their email confirmation process, a demographic questionnaire was automatically sent to their email address. Participants who completed the pretests received two reminders for the demographic questionnaire in the last month of data collection. Groups without rotation were not assigned any participants after a short duration to increase the number of participants in groups with rotation.

2.1.1.1. Piloting Process

All the tests and the training program were coded by the first author using recent web technologies such as HTML5 and KineticJS with cross-browser compatibility. The tests and the training task were tested on a variety of platforms such as Windows 7, Mac OS X, Ubuntu, tablet versions of iOS6 and Android 4.0.4 and on various browsers including Google Chrome, Internet Explorer 10, Mozilla Firefox and Safari.

All the tasks and the training program were initially intended for children of school age. The first piloting study was conducted with adults to reveal the major possible problems in our software and material. The second study was with children who used tablet computers in their daily educational setting between ages 8 and 13. Since we were obliged to conduct this study with a large number children between 15 and 30 due to the conditions and scheduling, children did not receive an appropriate testing environment. Apart from this, we did not have physical access to children and the compliance to the training was low. None of our child participants reached a satisfactory level in training attendance. Hence, we utilized this experience to develop a more user-friendly, understandable system with a better flow of the procedure with a large amount of feedback we received from children and teachers during this training attempt. Because all the tasks were online, and adaptable up to the adult levels (except for the online block design task), we decided to recruit adult participants to achieve higher compliance and more accurate testing results.

At the end of this process, in addition to the testing with typical children, we also piloted this study with two children with autism aged 8 and 11. These children were guided by a professional special education teacher who was trained and informed about the requirements of this study. Children fully understood the task requirements both in our tests and in the training. The older child, who was reported to be at a better cognitive level by the special education teacher, was assigned to the control group with rotation, which is less cognitively demanding compared to the experimental group, showed an increasing trend during the study. The younger one was assigned to the experimental group and this child showed a stagnating pattern during the study. As it is going to be discussed in the upcoming sections, considering that this child was younger and at the lower boundary of our typically developing age group and was exposed to a high cognitive demand, this result was a comprehensible outcome.

Overall, all these piloting studies assisted us to create a more user-friendly and better flowing online environment that we targeted adult trainees.

2.1.1.2. Demographic information

After the participants registered online and validated their email addresses, they received a demographic questionnaire (see Appendix A). One hundred and forty eight participants submitted the questionnaire. Sixty-one percent of them were female. The mean age of the responders was 28.28 and ranged between 16 and 58. The mean of the education they received in years was 16.17 and ranged between 10 and 25. Table 1 shows how often they played video games. Forty eight percent of the participants played computer games less than one hour a week or did not play at all. This showed that a large percentage of users were not highly devoted game players.

Table 1. *Percentages of frequency of video game playing.*

Frequency	**Percentage**
Does not play	32.4
Less than 1 hour a week	15.5
1-3 hours a week	24.3
3-5 hours a week	8.8
5-10 hours a week	10.1
10-20 hours a week	7.4
More than 20 hours a week	1.4

We also asked what kind of games they played in an open-ended question. Five raters who were knowledgeable about and experienced in video games were recruited online. We asked them to rate the genres or game titles in two dimensions. The first dimension was if at least one of the game titles or genres could be categorized as violent (e.g. Grand Theft Auto and Mortal Kombat) at least once and the second was whether they reported puzzle or quiz like games (Tetris, Candy Crush, Solitaire, 2048) including strategy games at least once. Table 2 presents the percentages rating of ratings for these two dimensions.

Table 2. *Percentages of ratings of experienced gamers in two dimensions.*

Rater	**Violent**	**Quiz-like**
1	23.81	86.90
2	30.95	77.38
3	22.62	75.00

4	36.59	91.03
5	43.02	83.72

As the table suggests, most users played quiz or puzzle like games that required developing strategies rather than responding in a fast and vigilant manner. Users who reported playing violent games did not constitute the majority of our participants.

2.1.2. Procedure

Before the training, participants took two sessions of pretests. In the first session, they took the forward digit span, the backward digit span, the block design, the letter span, the alphabet span and the arithmetic tests. This session approximately took between 40 and 50 minutes. In the second session, they took the Stroop task and the Standard Raven's progressive matrices test. This session took 30 to 40 minutes. After the second session of pretests, the participants were able to start the training task. The training phase was designed to take at least six weeks. They were instructed to join the training task 15 minutes a day (approximately 50 trials) and four times a week, which sums up to an hour of training for each week. Except for eight people, participants did not meet the suggested dosage of training. After the training procedure, they took the same tests again as posttests. They were invited to take these posttests even when they did not comply with the suggested training schedule. After the posttests, they were allowed to continue with the training task (the game program) as a token of appreciation for their participation. Thus, the shortest possible participation duration was 46 days, including pre- and posttest sessions. The participants received reminder emails every evening after 6PM in local time unless they chose to unsubscribe.

We developed eight online measurement tasks and a working memory training task (Gritty) with extensive testing over a period of a year including the testing of the software with diverse populations including typical children, two children with autism and young adults and middle-aged adults. We received large amounts of feedback to render the research program more accessible, understandable and easy to follow. Tests also revealed flaws in writing to the database under specific conditions (e.g., in case of losing connection with the server and time-out problems in user sessions) and these problems were debugged before the main study began.

2.2. Pre- and posttest measures

Forward digit span (FDS): In the forward digit span task, participants were presented one by one with a sequence of numbers from 1 to 9 in the middle of the screen and later asked to click on the correct numbers on the screen in the same sequence using a keypad. Each trial started with the presentation of a plus sign in the middle of the screen (see Figure 2).

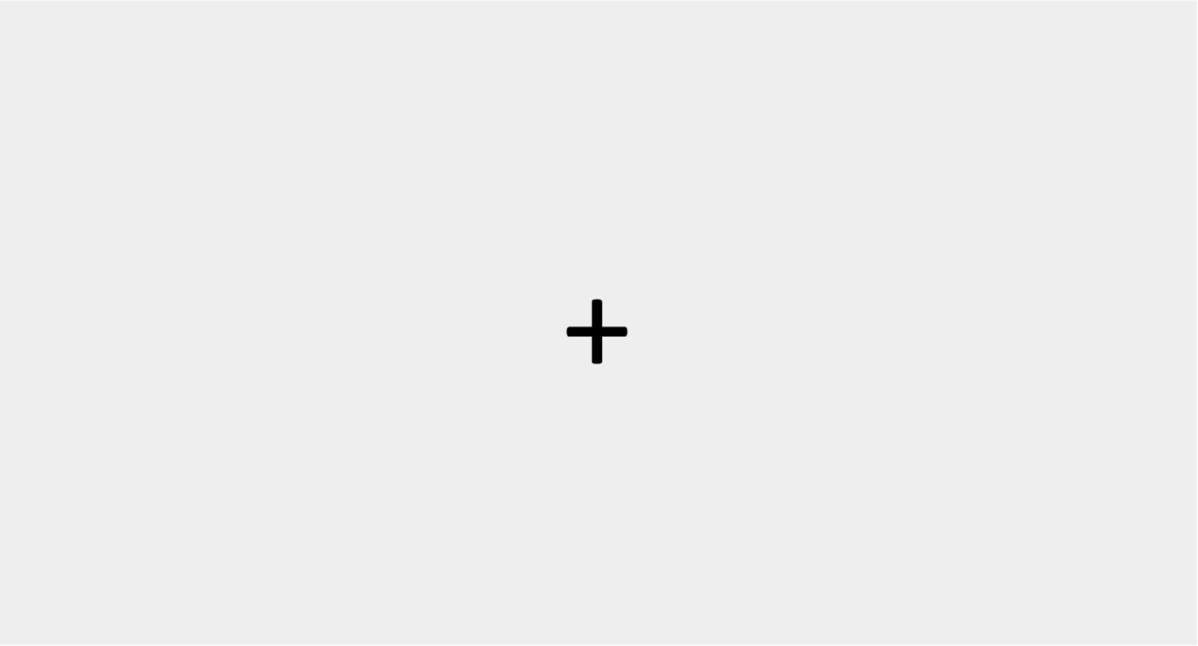


Figure 2. *Participants were shown a plus sign before the beginning of each trial in span measures to ensure that all participants start from the same area on the screen.*

Participants were asked to look at this plus sign for 500 milliseconds. Each number stayed for 600 milliseconds and then disappeared (see Figure 3).

Figure 3. *Span stimuli were shown in the middle of the screen.*

Later, the participant was asked to enter the numbers using a keypad in the order they were presented on the screen (see Figure 4).

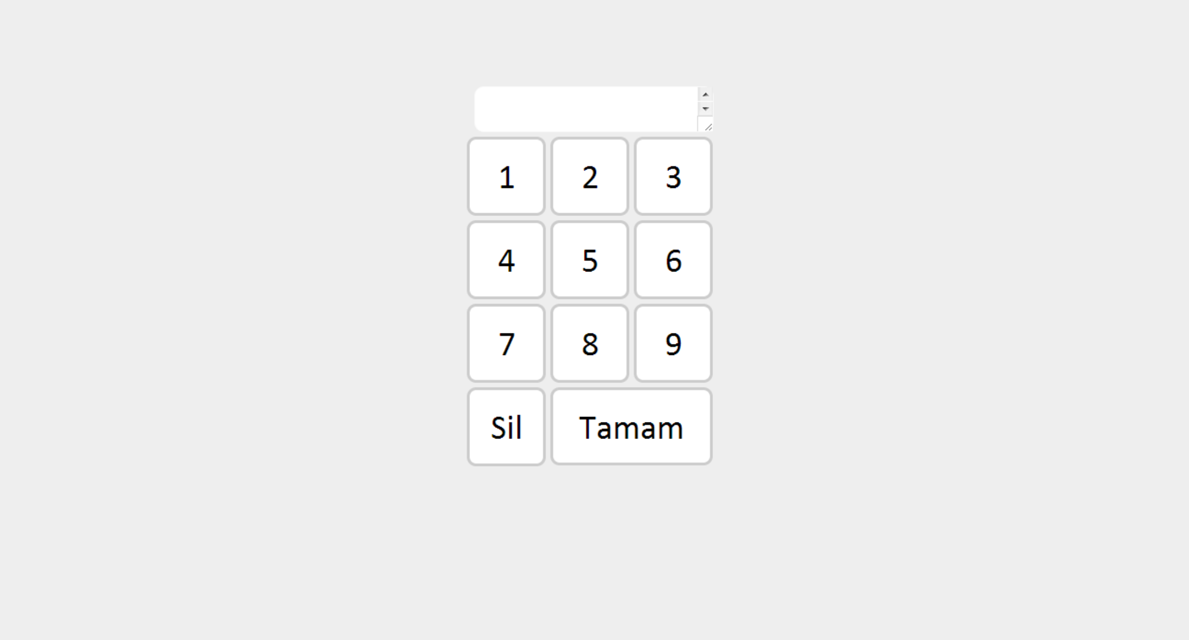


Figure 4. *Digit span responses were received with a special screen keypad.*

The numbers that the participant clicked stayed on the screen until the participant pressed the submit button. If the participant made three incorrect trials in a row, the task was discontinued. All participants started with three numbers in each trial. Each level consisted of three trials. If the participant could complete a level, the number of items in a sequence was increased by one. The highest level consisted of nine items. See appendix B for the material used in this task.

Backward digit span (BDS): This task was identical to the forward digit span except that participants were asked the numbers in reverse order. See appendix C for the material used in this task.

Letter span (LS): This task was identical to the forward digit span except that participants received a pseudo-random non-word sequence of letters. Each trial included a number of letters pseudo-randomly selected from 23 letters; that is all the English alphabet excluding Q, W, and X, since they do not appear in the Turkish alphabet. Answers were collected with the assistance of a special screen keyboard which includes only the letters in the task (see Figure 5).

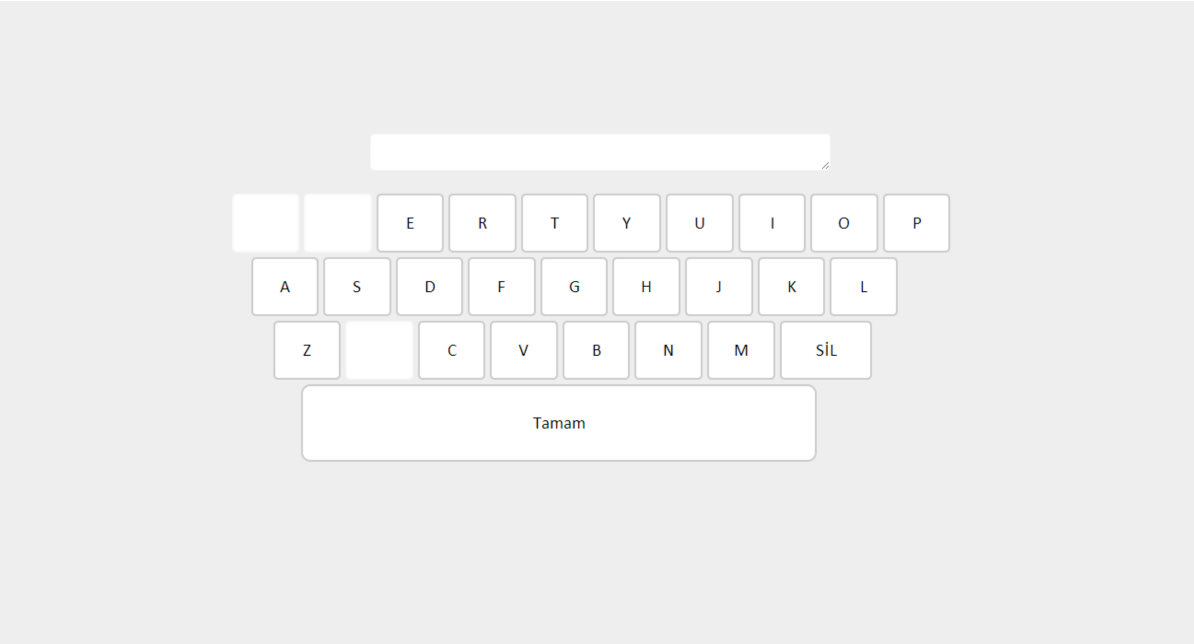


Figure 5. *Letter span and alphabet span tasks had a special screen keyboard composed of all letters used in the tasks.*

See appendix D for the material used in this task.

Alphabet span (AS): This task was identical to the letter span task except for that participants were asked to input the letters in alphabetical order. See appendix E for the material used in this task.

Block design: This task consisted of four boards on the screen: (1) the top ribbon, (2) the source board (on the left), (3) the target board (in the middle), and (4) help frame (on the right, see Figure 6).

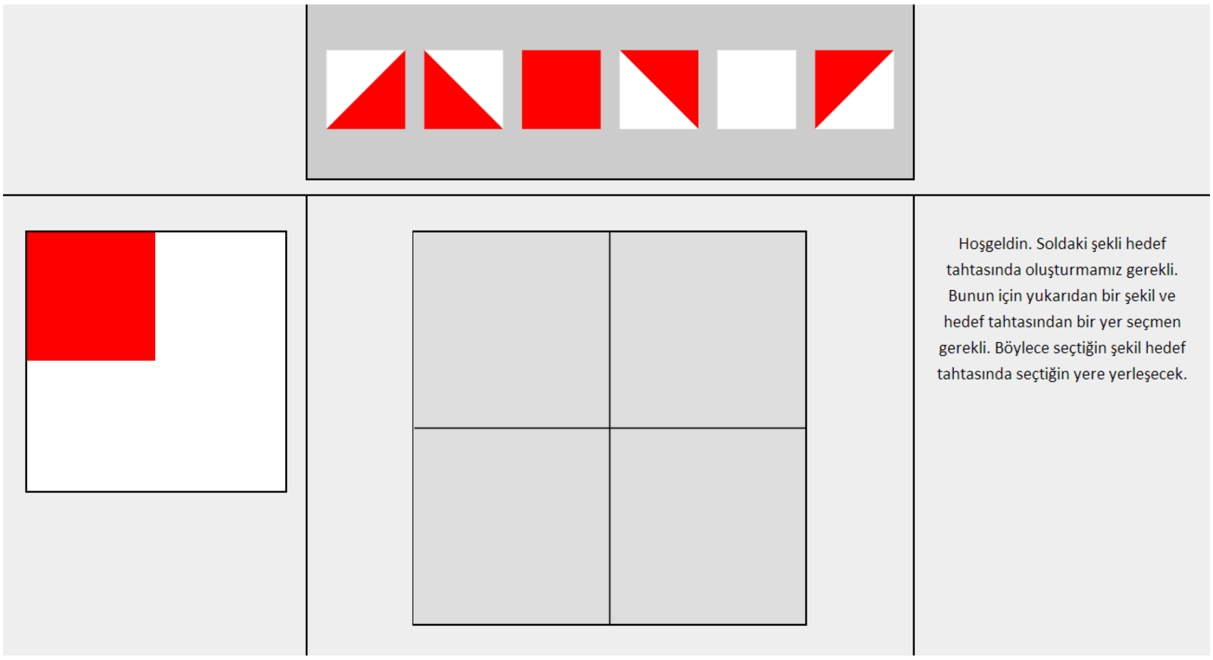


Figure 6. *Four areas of block design task: (1) the top ribbon, (2) the source board (left), (3) the target board (middle), and (4) the help frame (right).*

Participants were asked to build the pattern shown on the left (i.e., the source board) in the middle of the screen (i.e., the target board) using the blocks in the top ribbon.

The top ribbon contained six blocks to be used in the task. In this area, participants were presented a white square, a red square, and four white-and-red squares. These white-and-red squares were made up of two-equal sized red and white triangles. All possible right angle rotations were used in each trial (see Figure 6, the top ribbon). These six blocks were ordered randomly in each trial (see Appendix G for images used in this task).

The source board contained images made up of the blocks used in the top ribbon. It did not have grid lines that would allow parsing the image into blocks very easily. However, the target board always had grid lines to mark the areas that must be filled. Empty slots in the target board were marked with dark grey.

To copy a block onto the target board, participants were instructed to choose a block and select the position in the target board. If they first chose a target area, the area was highlighted in green and they were prompted in the help frame to choose a block to copy in this target area.

The first eight trials contained four blocks (two by two, block design 4) and the following six trials contained nine blocks (three by three, block design 9). Trials finished when all the blocks are placed correctly and the participant was congratulated. There was also a time-out feature to prevent participants from spending too much time on a single trial. The time-out duration was 90 seconds in trials with four blocks and 120 seconds in trials with nine blocks. When the time-out happened, the participant was prompted with a dialog box overlaid on the task indicating that they did not respond for a long time and they could continue with the next trial. We observed that the time-out duration was reasonable when the participants were actively engaged with the task. Trials that ended with a time-out were coded separately and as incorrect and were excluded from analyses. All moves in the target board were logged along with their time information.

Raven's standard progressive matrices (RSPM): RSPM is a widely used test as a measure of fluid intelligence and general nonverbal ability. This task was applied as described in the manual of the test (Court & Raven, 1995) except for it was presented in a computerized fashion. The offline version of the test was standardized for Turkish population by Düzen and colleagues (2007). Participants were presented with a visual pattern (see Figure 7).

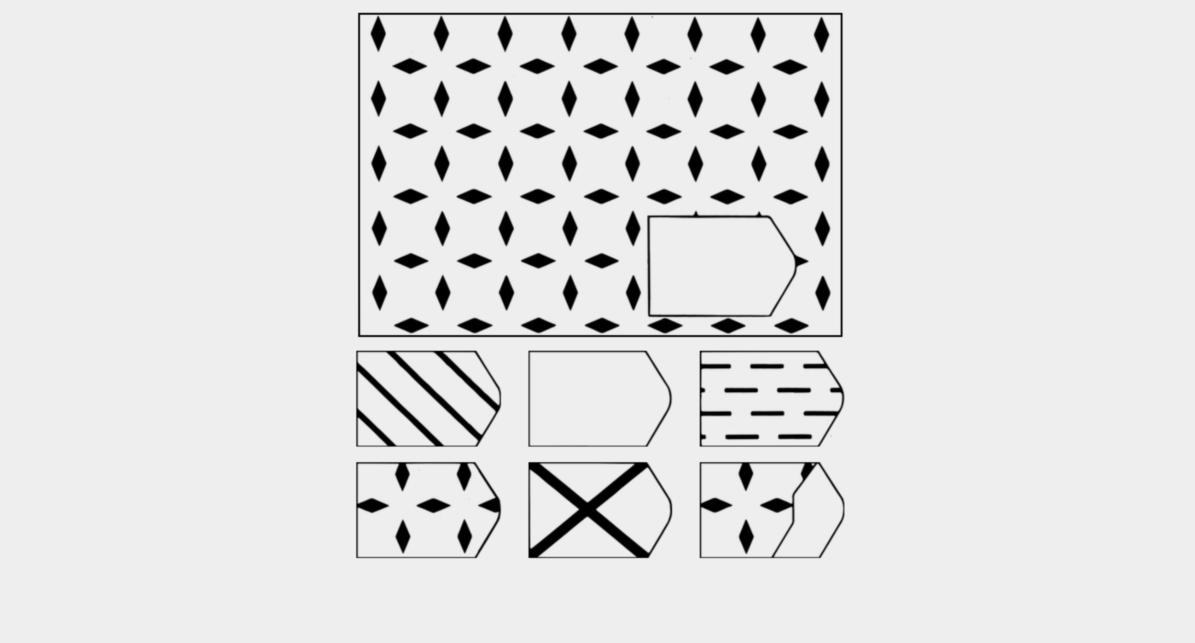


Figure 7. *A sample screen from Raven's Standard Progressive Matrices; the top and large pattern has a missing part. There are six or eight answer options at the bottom of the screen.*

The bottom right section of the pattern was missing. Possible answers were presented under the pattern. Participants were asked to click or tap on the correct pattern. This task contained five sets and each set contained 12 patterns to be completed. All sets and all patterns within a set were organized by their level of difficulty. Answers and reaction times were recorded. Upon the user response a new trial was presented. No feedback was provided.

Arithmetic: Participants were presented with 20 mathematical problems. Each problem required conducting one or more types of arithmetic operation: addition, subtraction, multiplication, and/or division. Problems had an increasing complexity. A keypad was shown under the problem and participants were asked to input their answer using these keys (see Figure 8).

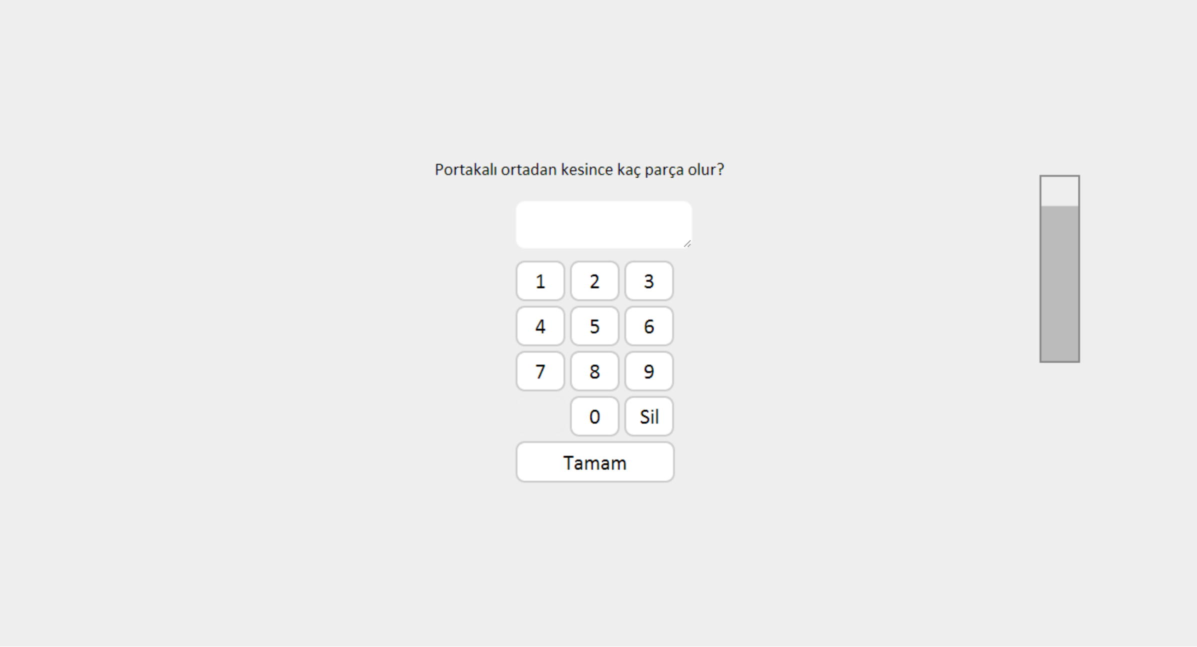


Figure 8. *A screenshot from the arithmetic task. The question and keypad are centered on the screen and there is a timing bar in the right side of the screen.*

Each question was timed. Time was shown by scaling the percentage into a fixed sized rectangle throughout the trials (see Figure 8, right). Time given for each question ranged between 75 and 120 seconds (see Appendix F for the questions and their durations). Dark grey area represented the percentage of the remaining time. When the allocated time finished, participants did not receive an extra prompt and they were allowed to input their answers. However, these answers were counted as incorrect responses. If they could not answer three questions in a row correctly, the task was discontinued. If they intended to input an empty answer, they were asked whether they would like to skip the question or spend more time on it. Skipped trials were also counted as incorrect trials.

Stroop: Participants were asked to indicate the color of the word or signs on the screen. There were four colors used in the task: red, green, blue and yellow. All stimuli were presented on a black background. There were three types of trials: (1) neutral trials where participants named the color of a pattern (i.e., XXXXXX), (2) congruent trials where the name of the color was written in the same color, and (3) incongruent trials where the color name and the color used to write it did not match. Participants received all possible incongruent trials (12 trials) once, all possible congruent trials three times (12 trials) and all possible naming trials three times. They responded by clicking one of four buttons placed at the bottom of the screen. Buttons were placed next to each other horizontally. Two buttons were placed at the left side and the remaining two were placed at the right side of the screen. The order of buttons was assigned randomly at each trial (see Figure 9).

Figure 9. *A sample screenshot from an incongruent trial in Stroop task. Four buttons at the bottom are the names of the colors as choice options. The text is "yellow" in Turkish but it is typed in green.*

All participants received a practice session before they started the task. The composition of the trials in the practice session was proportionate to the original task. All participants received 12 practice items. If the last six trials were all correct, they were able to receive the original task. If they had errors in the last six trials, they received new trials until when the last six trials were all correct. If they could not pass the practice in 24 trials in total, the task was ended. They received feedback only in practice trials. Their reaction time and accuracy were logged and used for analyses.

2.3. Training

2.3.1. Training Design

The main aim of the training was to improve the visuospatial short term storage and information manipulation skills of the participants.

The training task used a 3 × 3 grid plain (see the bottom square in Figure 10) and a ribbon at the top of the screen. The stimulus set consisted of nine animal images. These animals were bear, cat, chicken, cow, duck, fish, monkey, rabbit, and zebra. All of them were disyllabic in Turkish (see Appendix H for a list of Turkish names for the animals). They were among the animals that are learnt the earliest in children (Acarlar et al., 2009); hence, all of them were expected to be very familiar to participants. Images were designed in black and white to control for possible effects of color on memory.

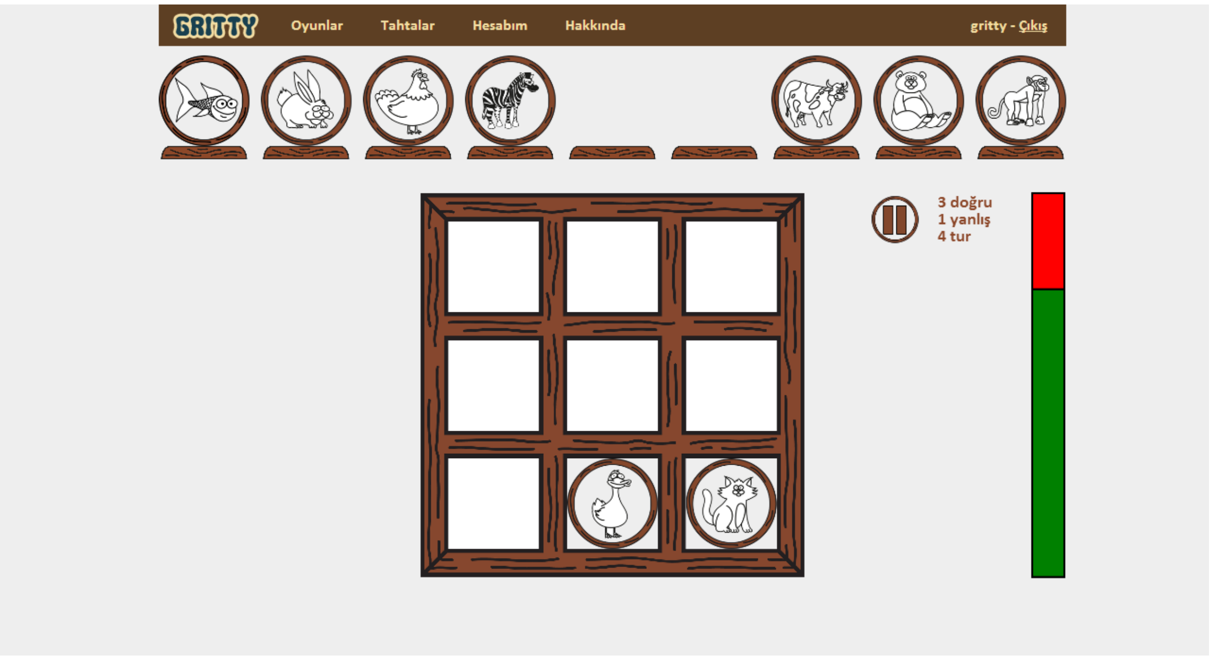


Figure 10. *A screenshot from the training task while waiting for the participant to respond (the top ribbon is open only when user input is allowed).*

The task started with presentation of a number of animal images in one of nine locations in the grid. All animal images were selected at random without replacement. Each animal was shown without removing the previous ones from the grid. The presentation duration of each new item was 1000 milliseconds. The order of presentation was from left to right and from top to down. Trials consisted of one to nine items and the number of the items presented in a trial defined the size and difficulty of a trial. After the presentation of the last item, all items stayed for 1000 milliseconds in the grid and later disappeared. Later, all nine animals were presented in the top ribbon with an animation from left to right (see the top ribbon in Figure 9). The order of the animals in the ribbon was randomized in each trial. The aim of the task was to find the items that were present in the trial and place them into the locations they were shown. Animal images in the ribbon were draggable. When an image was dragged into the grid, it was placed to the closest square in the grid. If the point at which the dragging ends was not in the grid, the image was placed back to its original location. Images previously placed in the grid could be removed and placed back to their original position by dragging them outside the grid.

Users had a button to pause and play the game when needed. This button allowed users to receive a new trial. Although technically this is not different than not responding for a long time, users reported that they felt more comfortable with the addition of this option. We also created a bar at the right side of the page that showed the percentage of correct and in incorrect answers in green and red respectively. On the left side of this bar, we placed a text box showing the total number of trials, total number correct trials and total number incorrect trials.

When the number of animals placed in the grid reached the number of animals presented in the trial, any user action was prohibited and a feedback phase began. All locations used in the initial presentation were superimposed with either a tick or cross sign. If the location had the correct image, it took a tick, otherwise a cross sign (see Figure 11). The feedback screen was presented for 1000 milliseconds.

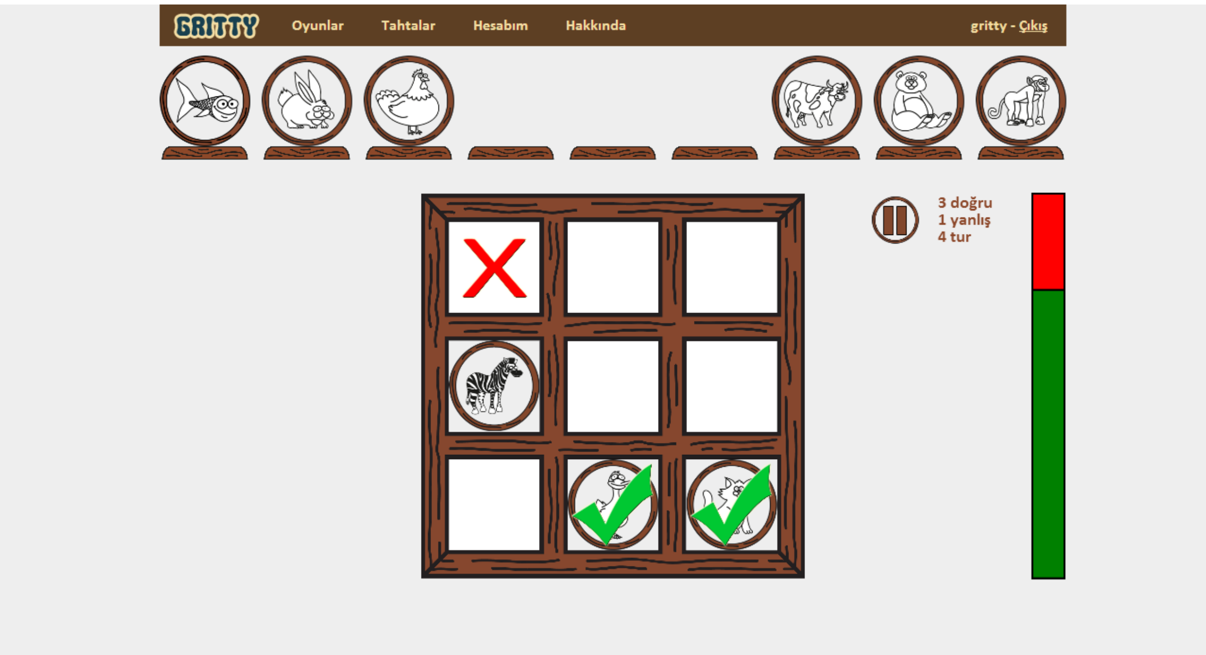


Figure 11. *Screenshot of a feedback phase in the training session.*

There were two intervention groups. The first group received a training targeting only the storage component of working memory and the second group trained for both manipulating and retaining information. The latter group had their grid rotated ninety degrees clockwise or counterclockwise (assigned randomly) in half of trials. Hence, they had to remember the position before the rotation and find the corresponding location after the rotation. Control groups received the same treatment except for the difficulty adjustment in the intervention groups. In difficulty adjustment, when the participants correctly completed three trials in a row in the same level, the difficulty of the game was increased by adding one more item to the trial until nine items were reached. If the participant failed to complete three trials in a row at this level, the number of the icons was decremented by one until there was only one item. This adjustment ensured that participants were always asked to perform at their highest skill level and pulled towards a higher level. Unlike these intervention groups, control groups received an increase in difficulty 30% of the time. Therefore, they spent more time reaching their level and in average they were pulled less towards levels that were above their current level.

2.3.2. Positive feedback

To increase the variability in practice, participants were able to share their progress in the game through a social feed that was placed in the website of the game (the boards section). This feedback for the game was given via two channels: (1) personal private pages that listed all the achievements gained in the training (left column), and (2) a public feed page (right column). When an achievement was gained the participant had the option to share or keep it private. The public feed page included all the feeds that were shared in the last week. Figure 12 shows the "boards" page for feedback information.

Figure 12. *A sample screenshot from the "boards" page displaying the badges, attendance information and shared achievements of other users. Names are hidden with rectangles for privacy of the participants.*

2.3.2.1. Types of feeds

There were two types of publishable feeds: badges and instant feeds.

2.3.2.1.1. Instant Feeds

Instant feeds were grouped into three categories: (1) playing the game for a number of trials, (2) being among the best players of the game, and (3) returning to the game after a while.

There were two feeds for the first category. Participants received an "ant" feed when they completed 50 trials. If they continued to play up to 100 trials they were awarded with "super ant" feed. The ant feed was as follows: "[The name of the user] has been gained the ant achievement!" The super ant feed was as follows: "[The name of the user] has gained the super ant achievement!"

The second category of instant feeds was based on calculations of the scores of the previous day. The first feed was given to three participants from each group. It was gained by having the best accuracy scores. This feed did not include the details of the achievement but it used an animal imagery: "[The name of the user] was a lion yesterday!" The second feed was given for the speed in the game. Excluding the three achievers, the fastest three participants were determined. Similar to the first feed in this category, the feed was as follows: "[The name of the user] was a cheetah yesterday!"

The third category included only one feed. This feed was for participants who did not join training for at least four days and later joined again. They were given this feed for their returning effort. It was worded as follows: “[The name of the user] is back with us and playing hard!”

2.3.2.1.2. Badges

Badges were graphical objects that were placed in the private page. In the beginning of the game all badges were gray scale with a lock image placed on the badge image. When the participant unlocked a badge, the graphical object gained color and the lock was removed. The training had a total number of ten badges that comes in two categories: (1) attendance badges, and (2) achievement badges.

2.3.2.1.2.1. Attendance Badges

The first badge, the first visit badge, was unlocked when the participant played the game for the first time. The second badge, the full week badge, was unlocked when the participant played the game five times a week (from Monday to Sunday evening). The following three attendance badges were unlocked with full week visits. The third badge was for two full weeks, the fourth badge was for four full weeks and the last badge was for six full weeks.

2.3.2.1.2.2. Achievement Badges

Achievement badges were unlocked based on the records in the second category of instant feeds. There were five achievement badges: The first two of them were for accuracy (i.e., lion) feeds. The first one was unlocked when the participants gained the accuracy feed twice and the second one was unlocked when the participant gained the accuracy feed four times. Another two of them were for speed feeds (i.e., cheetah). Similar to the first two, they were opened for two and four speed feeds. The last achievement badge opened when the participant gained one feed from the accuracy and one feed from speed.

CHAPTER 3

RESULTS

3.1. Data Preparation

3.1.1. Data Entry and Preparation for Cognitive Tests

Accuracy, reaction time, and input values in all the tasks were recorded in online database tables. The data tables for the cognitive tests included more than 300,000 rows and the one for the gaming data had more than 60,000 rows. All the data were visually screened before the analysis began to ensure the integrity of the collected data. Duplicate entries were cleaned.

Accuracy scores were created by counting the number of trials with accurate responses. For all the tasks, inaccurate trials were not included in the analysis. In some trials participants had extremely long reaction times. In these trials, probably they took a break. Reaction times greater than three standard deviations within an individual subject's data were excluded to ensure that trials with extreme reaction times and breaks were not in the analyses. For all the tasks, medians of reaction times of accurate trials were preferred because medians are more resistant to outliers. In the Stroop tasks, practice trials were not included into the analysis. In the block design task, trials that were finalized with a time-out were incorrect and excluded from the analysis; reaction times for trials with four items and nine items were calculated separately.

As a measure of working memory, we calculated difference scores by subtracting scores of BDS from FDS and scores of AS from LS, as the latter ones tapped into only STM while the former ones required manipulation of information by the participant.

3.1.2. Data Preparation for Training Scores and Formation of Contact and No-Contact Groups

Descriptive statistics such as minimum, maximum, counts, means, medians, skewness and kurtosis values were calculated for each participant for accuracy (the number of correctly placed items on the grid), level (the number of items that the participant was exposed to) and reaction time (completion time of a trial) for each day if they played at least one trial on that day. These values were recorded in database tables and visually screened. Means of four variables were calculated for analysis: mean accuracy, maximum accuracy, mean level, and maximum level. Participants who received less than 10 training sessions were grouped as no-contact group. This procedure was conducted to have a larger sample size and provide a comparison with previous studies which employed no-contact groups as the control group.

3.2. Analysis of Test Scores

3.2.1. Reliability of Test Scores

3.2.1.1. Reliability of Test Scores in the Whole Sample

Table 3 shows the Pearson correlations between pretest and posttest scores for each task. All tasks except accuracy in RSPM, RT in Block Design 9 and Stroop task had significant pretest-posttest correlations.

Table 3. *Test-retest reliability of task scores.*

	N	*r*	*p*
FDS (Accuracy)	47	.494	<.001
FDS (RT)	47	.502	<.001

BDS (Accuracy)	34	.614	<.001
BDS (RT)	34	.769	<.001
LS (Accuracy)	35	.517	.001
LS (RT)	35	.769	<.001
AS (Accuracy)	35	.609	<.001
AS (RT)	35	.495	.003
Arithmetic (Accuracy)	41	.44	.004
Arithmetic (RT)	41	.725	<.001
RSPM	23	-.033	.883
Block Design 4	24	.736	<.001
Block Design 9	24	.23	.28
Stroop Incongruent-Neutral	20	.163	.493
Stroop Incongruent-Congruent	20	-.243	.302

3.2.1.2. Reliability of Test Scores in the No-Contact Group

Because some of the participants received a substantial amount of training which may undermine the validity of results presented in the previous section, reliability analyses were also conducted in the no-contact group. Table 4 shows the correlations between pretest and posttest scores for each task in the no-contact group.

Table 4. *Test-retest reliability of task scores in the no-contact group.*

	N	*r*	*p*
FDS (Accuracy)	33	.567	.001
FDS (RT)	33	.439	.011
BDS (Accuracy)	19	.815	<.001
BDS (RT)	19	.612	.005
LS (Accuracy)	21	.384	.085
LS (RT)	21	.718	<.001
AS (Accuracy)	22	.699	<.001
AS (RT)	22	.361	.098
Arithmetic (Accuracy)	26	.479	.013
Arithmetic (RT)	26	618	.001
RSPM	13	-.112	.715
Block Design 4	19	.626	.004
Block Design 9	19	.28	.246
Stroop Incongruent-Neutral	11	.47	.144
Stroop Incongruent-Congruent	11	-.189	.578

We obtained the same pattern of correlations except for the marginally significant correlation in the accuracies of letter span. However, most of the correlations were lower than the whole sample. This result may be due to the lower sample size in this group.

3.2.2. Correlations among Test Scores

Table 5 shows Pearson correlations among test scores in pretests.

Table 5. *Pearson correlations among test scores in pretests.*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. FDS Accuracy														
2. FDS RT	-.108**													
3. BDS Accuracy	.566**	-.067*												
4. BDS RT	-.141**	.535**	-.069*											
5. LS Accuracy	.533**	-.094**	.536**	-.118**										
6. LS RT	-.130**	.679**	-.104**	.593**	-.166**									
7. AS Accuracy	.409**	-.058	.483**	-.058	.501**	-.097**								
8. AS RT	-.152**	.387**	-.068	.492**	-.165**	.569**	-.100**							
9. Arithmetic Accuracy	.167**	-.020	.206**	-.076*	.215**	-.130**	.212**	-.091**						
10. Arithmetic RT	-.197**	.416**	-.155**	.395**	-.209**	.548**	-.123**	.434**	-.149**					
11. RSPM	.135**	-.048	.137**	-.100*	.186**	-.041	.130**	-.096*	.134**	-.061				
12. Block Design 4	-.199**	.420**	-.231**	.466**	-.241**	.651**	-.203**	.457**	-.237**	.403**	-.038			
13. Block Design 9	-.171**	.346**	-.192**	.524**	-.221**	.459**	-.178**	.405**	-.122**	.339**	-.101*	.554**		
14. Stroop Incongruent-Neutral	-.055	.039	-.075	.032	-.058	-.010	-.079	.053	-.079*	-.007	-.053	.001	.032	
15. Stroop Incongruent-Congruent	-.051	.036	-.069	.030	-.053	-.009	-.072	.051	-.076	-.010	-.061	.001	.031	.997**

* $p < .05$, ** $p < .01$

These correlations showed that the memory tasks (FDS, BDS, LS and AS) were clustered together and tapped into similar dimensions. The Arithmetic test and the RSPM had significant but lower correlations with other tasks, as expected. The Block design had higher correlations with the BDS, AS, and the arithmetic task. This result is sensible given that they are both memory tasks that require manipulation of information. Stroop task did not correlate with the other tasks.

Because Pearson's correlation coefficient has assumptions of normal distribution and homoscedasticity and it is highly sensitive to outliers, we conducted the same analysis with a non-parametric correlation method, namely, Spearman's correlation. Although, this non-parametric analysis is more conservative compared to its parametric counterpart, the number of significant correlations has increased. Table 6 presents the Spearman correlations. Cells are color coded with green, red, blue and white. White cells are non-significant in both methods. Green indicates that the correlation was significant in both Spearman and Pearson methods. Red indicates that it was significant in the Pearson method but not in the Spearman method. Finally, the cells marked with blue shows correlations that are significant with the Spearman method but not in the Pearson method.

Table 6. *Spearman correlations among test scores in pretests.*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. FDS Accuracy														
2. FDS RT	-.128**													
3. BDS Accuracy	.573**	-.073*												
4. BDS RT	-.117**	.698**	-.041											
5. LS Accuracy	.542**	-.136**	.539**	-.098**										
6. LS RT	-.152**	.772**	-.082*	.656**	-.176**									
7. AS Accuracy	.412**	-.075*	.479**	-.015	.493**	-.082*								
8. AS RT	-.154**	.574**	-.085*	.623**	-.205**	.654**	-.139**							
9. Arithmetic Accuracy	.174**	.032	.226**	.011	.226**	-.025	.209**	-.056						
10. Arithmetic RT	-.192**	.455**	-.121**	.454**	-.211**	.523**	-.120**	.480**	-.119**					
11. RSPM	.110**	-.087*	.154**	-.106**	.190**	-.105**	.192**	-.135**	.118**	-.120**				
12. Block Design 4	-.205**	.589**	-.187**	.482**	-.259**	.617**	-.172**	.550**	-.153**	.433**	-.135**			
13. Block Design 9	-.193**	.613**	-.160**	.539**	-.208**	.632**	-.151**	.541**	-.038	.395**	-.186**	.632**		
14. Stroop Incongruent-Neutral	-.133**	.081	-.136**	.039	-.099*	-.008	-.102*	.015	-.052	.038	-.054	-.008	.025	
15. Stroop Incongruent-Congruent	-.092*	.039	-.081*	.020	-.028	-.001	-.049	.003	-.043	-.026	-.018	-.017	.000	.528**

* *p* < .05, ** *p* < .01

All the correlations that were significant in the Spearman method but not in the Pearson method were in the expected direction and strength. The most striking difference was in the Stroop task. Because the only measure in the Stroop task was RT differences between conditions, it showed low but negative correlations with accuracies in memory tasks (FDS, BDS, LS and AS). LS and AS did not have significant correlations with the incongruent-congruent difference. Considering that this difference is smaller compared to the difference with the neutral trials, which did not require any reading and were substantially different from the congruent trials in terms of cognitive load, non-significant correlations may reach to significance with a larger sample size.

The major finding in both methods was that accuracies and RT in memory tasks were positively correlated with in themselves and they showed low but significant correlations between each other. Block Design showed the same pattern of being correlated with information manipulation tasks (BDS, AS and arithmetic) in both methods. Again in both methods, RT in the arithmetic task had the highest correlations with the memory tasks while all correlations in arithmetic test accuracy were significant but low.

3.2.3. Principal Component Analysis for Test Scores

Due to the high number of variables we calculated from test data and the complex pattern of correlations we obtained, we also conducted a principal components analysis. This analysis aims to reveal underlying components in the data.

We first observed the scree plot in figure 13.

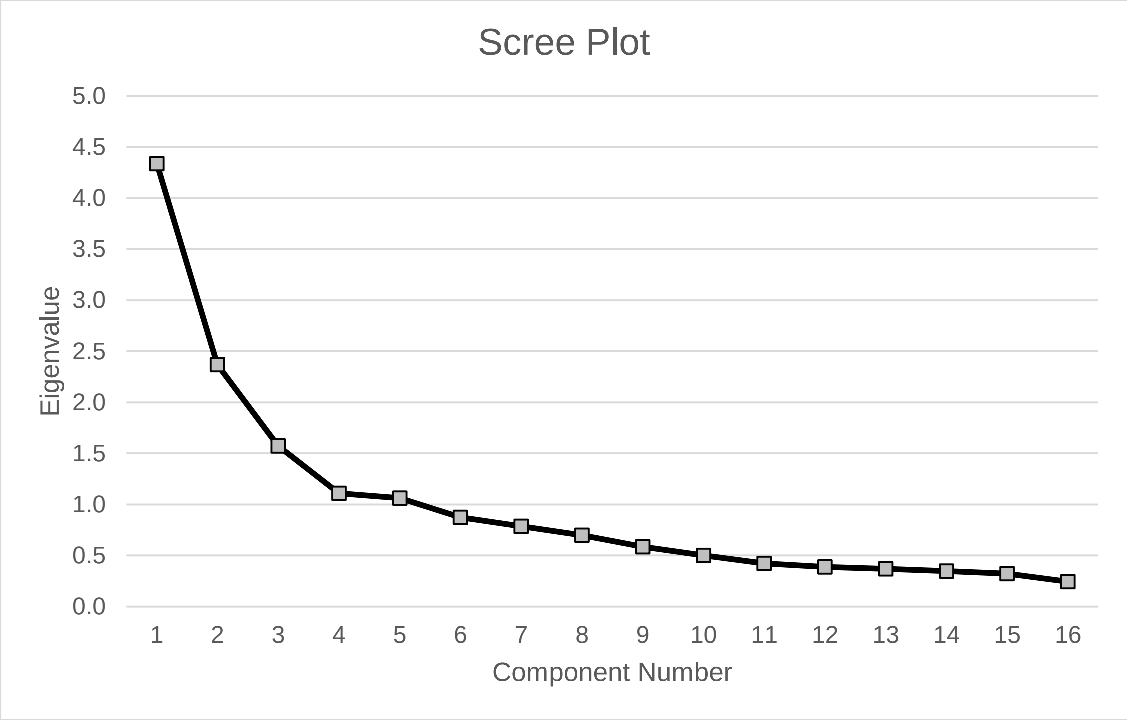


Figure 13. *The scree plot showing eigenvalues for each component number.*

There was only a minimal change from four components (with an eigenvalue of 1.11) to five (with an eigenvalue of 1.06). Hence, we decided that the critical elbow in the plot was on four components. Principal component analysis (henceforth, PCA) was conducted with four components and without a rotation (as opposed to the common oblimin rotation used in maximum likelihood analysis used in factor analytic analysis which requires large samples).

Table 7 presents the component loadings for each variable analyzed in correlation analyses. Component loadings greater than .4 are marked with a grey fill to visually present the loading pattern.

Table 7. *Component loadings for pretest variables.*

	Component			
	1	2	3	4
1. FDS Accuracy	-.454	.628	.042	-.255
2. FDS RT	.442	.318	.059	.431

3. BDS Accuracy	-.472	.674	.032	-.177
4. BDS RT	.714	.387	.098	.046
5. LS Accuracy	-.541	.606	.135	-.065
6. LS RT	.769	.420	-.082	.094
7. AS Accuracy	-.449	.604	.120	-.103
8. AS RT	.687	.284	.020	.058
9. Arithmetic Accuracy	-.225	.289	.151	.627
10. Arithmetic RT	.537	.141	-.215	.000
11. RSPM	-.187	.177	.014	.493
12. Block Design 4	.757	.168	-.101	-.233
13. Block Design 9	.715	.258	-.047	-.131
14. Stroop Incongruent-Neutral	.278	-.138	.824	-.066
15. Stroop Incongruent-Congruent	.218	-.090	.855	-.126

The analysis results revealed a clear pattern of component loadings. As the table suggested, the first component was about the RT performance of the participant. RTs in all the memory tasks, the arithmetic task, and block design scores showed positive and high component loadings to the first component. This meant that the most prominent component (i.e., associated with the highest variance) we found in the pretest data was the RT performance. Another striking pattern was found with the negative correlations of accuracies of memory tasks. There was no other task that showed smaller than -.4 loading with this dimension. Therefore, this dimension can be titled as the memory and problem solving RT of the participant. That is, even if it was also relied on the RT, it tapped into a separate domain.

The second dimension had high component loadings only from the accuracies in memory tasks. This dimension can be evaluated as the accuracy of overall memory performance of the participant.

The third dimension was the performance in the Stroop task, which is a typical inhibition task. Although this was also a measure in RT, it was dissociated from other measures. The fact that it had no other high component loadings, showed that it was not related to the other tasks.

As for the fourth dimension, the two greatest component loadings were received from accuracy in the arithmetic task and RSPM. Considering that accuracy in the arithmetic task (rather verbal) and in the RSPM task (rather visual) aimed to measure the problem solving skills of the participants, this dimension can be labeled as visual and verbal problem solving capacity.

3.3. The Nature of the Data and Understanding Attrition

The main problem during the data collection was the high number of dropouts. Although more than two thousand people signed up and more than a thousand started to take the pretests, many users did not finish the study or comply with the suggested schedule of training.

Because the initial attrition rate was extremely high, we discontinued assigning people to groups without rotation. Only 81 participants (4%) were randomly assigned to these groups and only two people reached up to 24 sessions.

3.3.1. Group Assignment and Attrition

Participants who played for 10 or more days were grouped as those who attended the training. Six participants in the experimental group and 17 participants in the control group satisfied this criterion although the initial numbers of participants in each group were very close (959 to 962 respectively). Chi-square analysis showed that there were more participants in the control group ($\chi^2(1, N = 1921) = 5.29, p = .21$) compared to the experimental group at the end of the study.

3.3.2. Predicting Attrition from Pretest Scores

Four hundred ninety seven participants reached to the training phase. We conducted multiple regression analyses and identified two variables as the meaningful predictors of number of sessions they received. RSPM scores ($\beta = .118, t(494) = 2.695, p = .007$) and median reaction times in the block design with four blocks ($\beta = .216, t(494) = 4.954, p < .001$) predicted the number of sessions ($R^2 = .06, F(2, 497) = 15.45, p < .001$).

3.4. Training Scores

3.4.1. Correlations among Training Scores

We analyzed inter-correlations among maximum accuracy, mean accuracy, maximum number of levels and mean number of levels scores across training sessions of those participants who attended at least five sessions. As the table 8 suggests, the correlations of measures in the training sessions were very high. We chose the maximum level the participant reached in each session to employ in the further analyses, because it is the best indicator of maximum working memory capacity at a given time in this training environment.

Table 8. *Correlations of measures of training performance.*

	1	2	3
1. Maximum Accuracy			

2. Mean Accuracy	.971**		
3. Maximum Level	.997**	.971**	
4. Mean Level	.916**	.959**	.930**

** $p < .01$

3.4.2. Training Scores over Time

Figure 14 and figure 15 present maximum levels in each training session for experimental and control groups respectively. All participants had steady or increasing trend lines. Scores showed a high fluctuation and augmentation until the fifth session. After the fifth session, they were mostly stabilized.

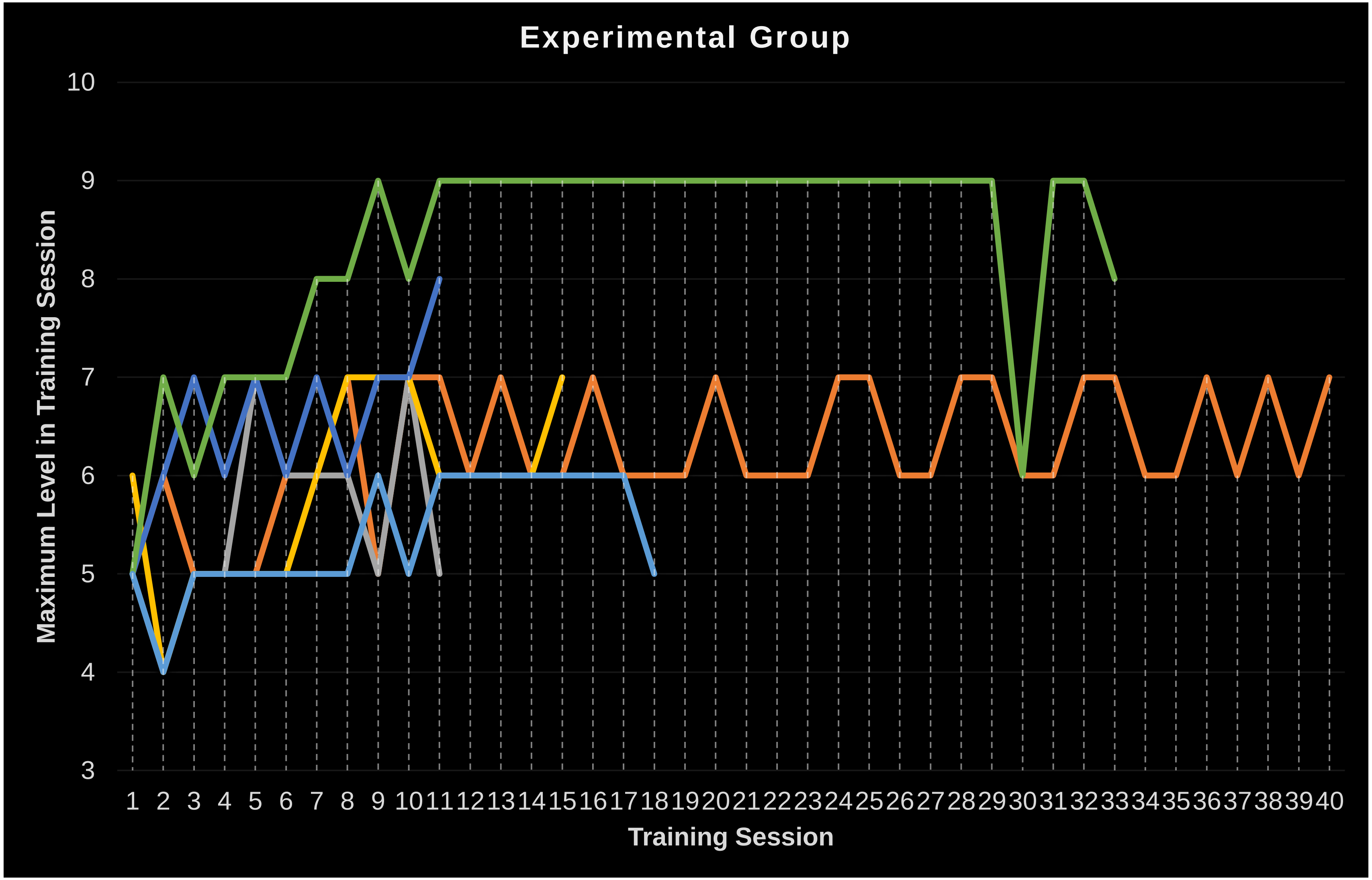


Figure 14. *Maximum level in training session in each session for 6 participants in the experimental group.*

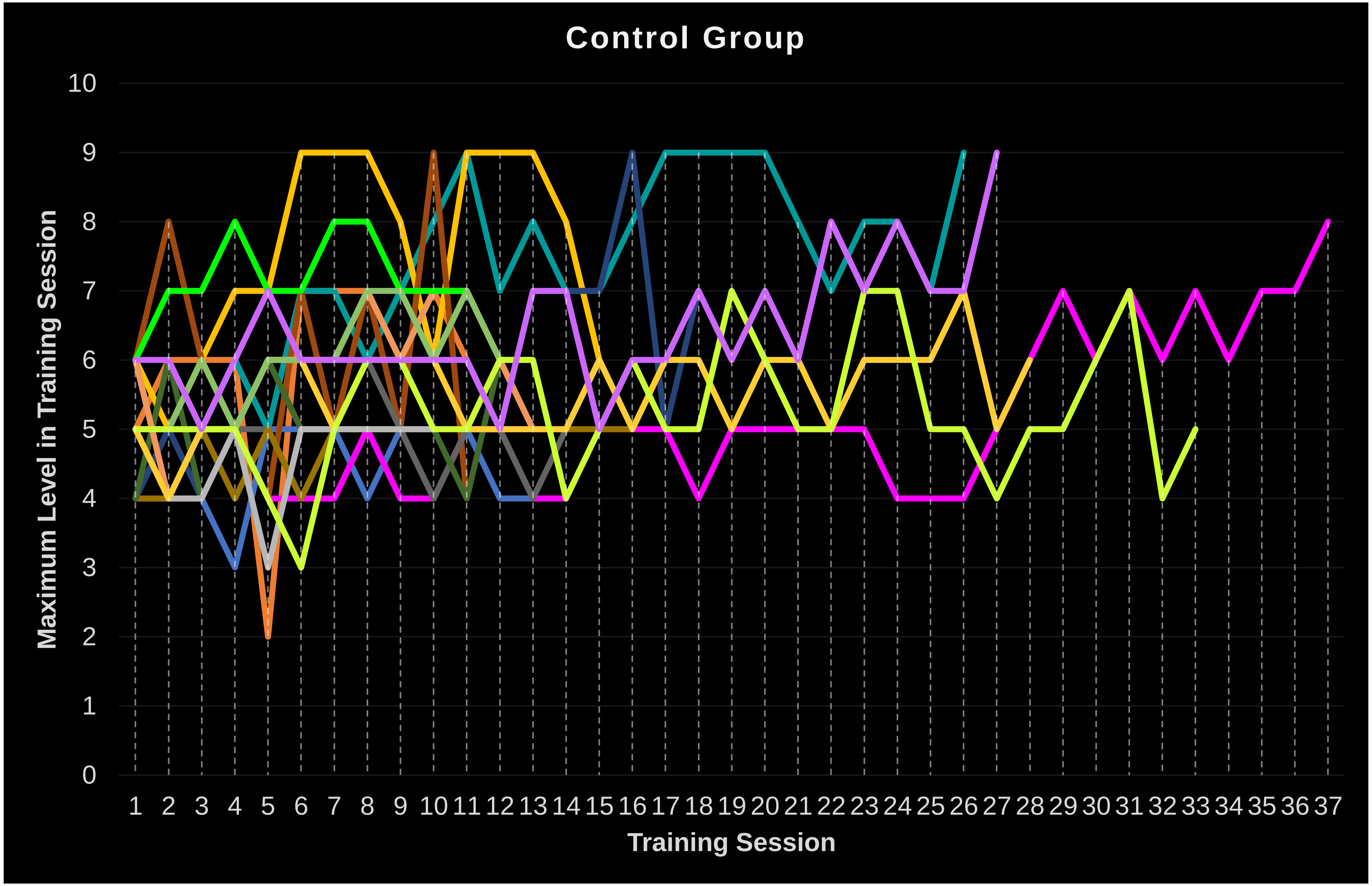


Figure 15. *Maximum level in training session in each session for 17 participants in the control group.*

3.5. Correlations among Test Scores and Training Scores

Table 9 shows correlations among test scores and training scores for participants who attended at least ten sessions (contact group).

Table 9. *Correlations among test scores and training scores for participants who attended at least ten sessions.*

	Maximum Accuracy	Mean Accuracy	Maximum Level	Mean Level
1 FDS Accuracy	.278	.240	.307	.255
2 FDS RT	.086	.052	.076	-.027
3 BDS Accuracy	.424*	.361	.425*	.297
4 BDS RT	-.105	-.186	-.127	-.231
5 LS Accuracy	.278	.282	.298	.250
6 LS RT	-.261	-.313	-.249	-.295
7 AS Accuracy	.346	.276	.356	.266
8 AS RT	.098	.130	.107	.045
9 Arithmetic Accuracy	.419*	.439*	.430*	.431*
10 Arithmetic RT	-.244	-.300	-.256	-.348
11 RSPM	-.249	-.130	-.228	-.042
12 Block Design 4	-.320	-.398	-.309	-.333
13 Block Design 9	-.217	-.295	-.215	-.268

14 Stroop Incongruent-Neutral	.078	.097	.086	.090
15 Stroop Incongruent-Congruent	.151	.168	.170	.221

* $p < .05$

Arithmetic accuracy and accuracy in BDS correlated significantly with all four training measures.

6. Group Comparisons

The first hypothesis was that training groups would have a greater increase in FDS, BDS, LS, AS and Stroop scores compared to the control groups. Due to low and probably uneven number of participants who completed the training and acquired valid tests scores, comparisons between experimental and control groups were not significant. To increase the sample size, we also asked participants who did not attend as regimented to the training session to complete the posttests. To analyze the differences between those who attended and those who did not, we merged the experimental and the control groups and compared them to the no-contact group (those who received no training or less than ten sessions).

We compared differences in digit and letter spans for contact and no-contact groups in pre- and posttests.

In digit spans, there was no difference between contact and no-contact groups and there was no difference between pretests and posttests. As expected, there was a significant interaction effect between testing time and group ($F(1,30) = 11.79$, $p = .002$, $MS_e = 8.94$, $\eta_p^2 =$

.28). This interaction effect showed the gap became smaller in the contact group, while it became larger in the no-contact group. Table 10 shows group sizes, means and SDs.

Table 10. *Sample sizes, means and standard deviations for differences of digit spans in pre- and posttests.*

	Contact		No-contact	
N	14		18	
	Pretest	Posttest	Pretest	Posttest
Mean	-2.29	.5	-.111	-2.5
SD	3.36	1.99	3.6	3.59

In letter spans, there was no difference between contact and no-contact groups and there was no difference between pretests and posttests. The interaction effect between testing time and group was marginally significant and the direction was towards a greater increase in the contact group compared to the no-contact group ($F(1,28) = 3.17$, $p = .086$, $MS_e = 12.35$, $\eta_p^2 = .10$). Table 11 shows group sizes, means and SDs.

Table 11. *Sample sizes, means and standard deviations for differences of letter spans in pre- and posttests.*

	Contact		No-contact	
N	13		17	
	Pretest	Posttest	Pretest	Posttest

Mean	-6.08	-4.23	-3.7	-5.12
SD	2.78	4.04	4.14	4.6

We failed to find any significant differences between the no-contact group and contact group in other test scores.

The second, third and fourth hypotheses required comparing groups with and without rotation. Since the groups without rotation were discontinued during data collection, we were not able to test this hypothesis. The fifth and sixth hypotheses required showing benefits of training by comparing experimental and control groups. Since we did not detect an effect, these hypotheses were not suitable for testing. The same situation is also valid for the effects of dosage.

CHAPTER 4

DISCUSSION

Since all working memory training studies to our knowledge were conducted in laboratories, the next step in the field was to carry the training outside the laboratory (i.e., in real life), the ultimate place where the training aimed to be conducted, and the main aim of this study was to test the feasibility of the online working memory training. This aim required two main steps to be realized: (1) creating online working memory tasks in line with the specific hypotheses we formulated and problem solving tasks that are related to working memory (block design, arithmetic, RSPM and Stroop) to test the transfer effects, and (2) creating a single original and substantiated working memory training task that will require information retention only or synchronous retention and manipulation of information.

4.1. Reliability of the Test Scores

We first studied the reliability of pre- and posttest scores. Most tasks we developed had good test-retest correlations. All the core working memory tests, the arithmetic task and block design 4 received satisfactory correlations. Correlations were not sufficient in block design 9, Stroop and RSPM. In the FDS, BDS and AS tasks, accuracies had higher test-retest correlations compared to the RTs.

The interesting point is observing that block design 4 had high test-retest correlations while block design 9 did not. There may be several reasons for this result. Firstly, the processing capacity of the majority of participants might have been insufficient to handle nine items simultaneously. Since the main measure in this task was RT of the participant during the whole trial, this task may be measuring only the fastness of the reaction of the participant but not the memory capacity. Secondly, they may be overwhelmed by the task requirements in trials with nine items.

In the Stroop task, the main measure was again only the RT and this task requires very low RTs unlike other tasks which require long RTs for data entry. Although we collected the data using the date-time information on the user's computer using JavaScript, when the Internet connection has a long waiting time for the server response, users may have been busy with other daily tasks or have had low motivation to respond fast. One should also consider that they may be using different hardware and/or software during pre- and posttests, which may undermine the accuracy of the results given that this is the fastest task.

As for the RSPM, which was the last task in the second session, our feedback from the participants was that they were overwhelmed by the length of the task. As the sample sizes were low (highest sample size being eleven), most users discontinued the task and the final sample size for this task was extremely low. These conditions may be responsible for non-significant correlations.

4.2. Correlations among Test Scores

In line with the findings in the test-retest reliabilities, clustering of memory tasks was better compared to the clustering of RTs. Accuracies and RTs had positive correlations among themselves and RTs and accuracies had low and negative correlations with each other. These findings showed that these tasks tapped into similar cognitive skills.

As for the arithmetic task, accuracies had low and significant correlations with accuracies in memory tasks. However, the correlations among RTs of memory tasks and arithmetic task were higher compared to the accuracies. These results showed that RT in arithmetic task was more susceptible to memory capacity compared to relation with accuracy in this task.

RSPM had low but significant correlations with several tasks. Contrary to our expectations, it did not have a strong relation with other problem solving tasks in Pearson correlations, this may be due to its low test-retest reliability and low sample size. At the same time, despite the low sample size we had, in the PCA, RSPM scores clustered satisfactorily with arithmetic task accuracy and not with other tasks. Although these tasks were problem-solving tasks in separate domains, their clustering showed that problem-solving component was a separate dimension.

The block design task in both levels (four and nine) showed the strongest relationship with RTs of BDS and AS. Since the only output of the task is the RT of the user and requires high levels of information manipulation, this finding shows that block design task can be used as a non-verbal (visual) problem solving task, which is closely related to RT in information manipulation. According to the PCA, the block design scores in both levels, had the highest component loadings to RTs in memory tasks but not with problem solving tasks. Both of these indicators points to the possibility that block design did not put high demands on problem solving capacity but the fast RT of the participant in memory tasks and not in the Stroop task.

Stroop task in general was not related to other task scores. This may again be due to its low test-retest reliability. As for the PCA results, we observed that both performance measures of Stroop clustered separately from all the other tasks. This finding again supported that memory tasks were separate from this inhibition task and tapped into a cognitive separate dimension.

If the correlations were too high, this would mean that they did not provide any extra information and if they were too low this would mean that skills that were supposed to have a common core were not measured in the appropriate way. To sum up, PCA showed that theoretically close measures clustered well with each other.

Overall, these correlations showed that memory tasks, arithmetic task and block design task were reliable and useful measures since in general they received significant correlations in appropriate levels. For memory tasks, considering their test-retest reliabilities and correlations among themselves, we showed that accuracies were better measures compared to RTs. Considering that arithmetic task had sufficient test-retest reliability and lower but significant correlations compared to the ones among memory tasks (FDS, BDS, LS and AS), it can be viewed as an online measure that merits further study.

4.3. Training Scores

In the first place, we studied the changes in training scores over time. Our graphs indicated that scores tended to either increase or remain stable over time although in most cases the trend was toward an increase. We also observed that when there was a trend toward increase, after the fifth session, the trend turned more into stabilization. Nevertheless, this does not mean that the optimum duration can be only five sessions, because participants may need more practice to internalize the changes and the practice effect that they experienced.

We next studied the relationship between training scores with pretest scores. As indicated in the Results section, maximum level was chosen as the main measure for the working memory capacity in each day's session. To receive more accurate results, we selected participants who completed at least ten sessions. The results showed that there were significant correlations between the mean of maximum level training scores and accuracy in BDS and arithmetic scores. This showed that the training tapped on manipulation of information in working memory, which is considered to be near transfer, and problem solving which is considered as far transfer. Being related to arithmetic scores rendered this training more promising for more immediate tasks for real life requirements. At the same time, having medium level correlations with BDS and not FDS showed that the task requires both retention and manipulation of information. However, this new task required manipulation of information in the visual domain and not the verbal domain. Hence, this new task could be considered to be promising both for measurement and training of working memory in the visual domain.

4.4. Effects of Training on Test Scores

Due to low compliance of participants and low sample size, analyses comparing control and experimental groups did not yield significant results. Considering that many studies were published even in high-impact journals employed the same method, we decided to compare participants who had training and who did not. To be more conservative, although we observed stabilization after the fifth session, we opted for using participants who had at least ten sessions. To test the effect on only in information manipulation but not retention we used difference scores between BDS and FDS and between AS and LS. There was a significant effect of training between these groups in digit span and a marginally significant effect in letter-based spans (AS and LS). This showed we found positive change in the contact group while we failed to find a difference between adaptive and non-adaptive training. These results also indicated that the training we conducted in the visual domain may be transferred into the verbal domain when contact and no-contact groups were employed but not when comparing adaptive and non-adaptive training. However, this result has to be evaluated considering smaller sample size in the latter analyses. In larger samples, this effect still may a have chance to be detected.

To sum up, we firstly showed that the FDS, BDS, LS, AS, block design 4 and the arithmetic task received satisfactory test-retest reliabilities. Moreover, all the tests in this study showed appropriate clustering in the principal component analysis. As for the training, in both groups we observed stabilization in maximum level session scores after the fifth session. Training task had moderate correlations with BDS accuracy that measures storage and manipulation in working memory and arithmetic RT which is a measure of fastness of response in verbal problem solving. There was no difference between the experimental and control groups while we detected a significant difference between the contact and no-contact groups in digit span differences and a marginally significant difference in letter span differences. Both of them were in the expected and desired direction.

4.5. Limitations and Future Directions

During the data collection period, the main challenge we had was keeping participants in the training program. The whole training program duration including pre- and posttests was 46 days. During the training period, although the total time of training in a week was only one hour, we asked participants to receive training at least four times a week. According to a large number of feedback messages from our participants, because of this heavy demand for attendance, we had a high attrition rate. Another reason for this attrition rate is that participants did not receive an immediate compensation for their effort, like cash payment. Our compensation was only one iPad Mini that we could award with a raffle and providing the test scores to the participant due to the limited budget of the study. Future studies may obtain smaller attrition rates if they can provide direct cash payment to all the participants for each session.

Unlike previous studies mentioned in the first chapter, we did not have physical access to the participants and control over their daily schedules, as researchers typically have in child training studies. Although we emailed subscribed participants reminders to continue the training and provided online game achievements (badges and motivational messages), the success of continuing the training relied only on the motivation of the participant.

Another challenge we encountered was the selective attrition of participants. The task in the control group was not designed to push the limits of working memory capacity of the participant whereas in the experimental group it did. Hence, participants in the experimental group were under a higher demand and cognitive load. Our results showed that, participants in the control group tended to attend more to the training sessions. Because of this, we did not have a large sample size in the experimental group and group sample sizes were disproportionate, which is not a desirable situation in statistical analyses.

We believe that future studies in online working memory training should provide cash compensation to the participants to test the effectiveness of the training program. Another step that has to be taken is providing more and better online achievements to the participants, such as providing social media shares on popular sites like Facebook, Google+ and Twitter. Utilization of social media may help to boost both the number of initial registrations and attendance by increasing the motivation of users. To have more control on the study, we also kept our graphics mostly black and white in images and simple and sleek elsewhere. Our feedback showed that participants ask for more catchy graphics that provide the feeling of playing a game rather than taking a psychological test. Another way to enhance the attendance motivation may be giving highly enjoyable filler tasks from time to time during the training to refresh the participants' attention and motivation. Due to technical limitations, our training was compatible with only tablet computer and not mobile phones, because this would create a less controlled environment. Considering that the share of mobile phones is in an incredible increase, mobile support is indispensable in future attempts to increase the participant compliance and attendance.

Despite the major challenges during the study, we achieved to develop new potential online measures and a training program. The measures that we developed are to be validated and tested with previous reliable and valid measures.

The best option for the validation of the memory tasks is to test the relationship of these measures with established and venerable computerized offline tasks such as operation span and sentence span. These tasks require both storage and manipulation of information. To our knowledge, there is no online task that measures these constructs. One option is to administer to FDS, BDS, AS, LS and the operation and sentence span tasks in the laboratory. However, given that the conditions cannot be equal to the conditions our participants who remotely took our online tasks, the best way for validation is to develop and test the online versions of these two established tasks in the first phase. In the second phase, the validity of these memory tasks can be tested with an online study. These validation processes are also valid for other tasks developed in this study.

Online block design test, which was built for this study for the first time, also requires further scrutiny. This task clustered mostly with RT measures of memory tasks and did not cluster with problem solving tasks. Given that the offline version of this task was intended for children, visual parsing may have been too easy for our adult participants. One option to test this is to provide harder and larger visual parsing problems to adult participants in a more controlled environment, such as in an offline fashion in the laboratory.

The greatest strength of this study was raising more questions and alternative ideas to be tested than providing conclusive answers. We believe that this first attempt of online measurement and training of working memory outside the confined laboratory environment is going to help researchers to study our tasks under development and their effects in a variety of participant populations such as children, adults and people with atypical cognitive development, under various conditions with virtually limitless opportunities of manipulation of structure, order and parameters in both testing and training.

REFERENCES

Acarlar, F., Aksu-Koç, A., Küntay, A. C., Maviş, İ., Sofu, H., Topbaş, S., & Turan, F. (2009). Adapting MB-CDI to Turkish: The first phase. In S. Ay, Ö. Aydın, İ. Ergenç, S. Gökmen, S. İşsever, & D. Peçenel (Eds.), *Essays on Turkish Linguistics: Proceedings of the 14th International Conference on Turkish Linguistics, August 6-8, 2008*. Wiesbaden, Germany: Harrassowitz Verlag.

Alloway, T. P., & Parker, T. E. (2011). *Training Your Brain for Dummies*. West Sussex, England: John Wiley & Sons, Ltd.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The Psychology of Learning and Motivation (Volume 2)* (pp. 89–195). New York: Academic Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11058819

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29. doi:10.1146/annurev-psych-120710-100422

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation (Volume 8)* (pp. 47–89). New York: Academic Press.

Bergman Nutley, S., Söderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: A controlled, randomized study. *Developmental Science*, *14*(3), 591–601. doi:10.1111/j.1467-7687.2010.01022.x

Bryck, R. L., & Fisher, P. a. (2012). Training the brain: Practical applications of neural plasticity from the intersection of cognitive neuroscience, developmental psychology, and prevention science. *The American Psychologist*, *67*(2), 87–100. doi:10.1037/a0024657

Buschkuehl, M., & Jaeggi, S. M. (2010). Improving intelligence: A literature review. *Swiss Medical Weekly*, *140*(19-20), 266–72. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20349365

Colom, R., Flores-Mendoza, C., & Rebollo, I. (2003). Working memory and intelligence. *Personality and Individual Differences*, *34*(1), 33–39. doi:10.1016/S0191-8869(02)00023-5

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by g, *32*, 277–296. doi:10.1016/j.intell.2003.12.002

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: the importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*(2), 331–335. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11495122

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163–183.

Court, J. H., & Raven, J. (1995). Manual for Raven's progressive matrices and vocabulary scales. In *Section 7: Research and References: Summaries of Normative, Reliability, and Validity Studies and References to All Sections*. . San Antonio, TX: Harcourt Assessment.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Models of Active Maintenance and Executive Control* (pp. 62–101). Cambridge: Cambridge University Press.

Cowan, N. (2005). *Working Memory Capacity: Essays in Cognitive Psychology*. East Sussex, UK: Taylor & Francis Group.

Dobrzenski, B. A. (2009). *Cognitive Consequences of Working Memory Training on a Typically Developed Population*. University of Notre Dame, Notre Dame, Indiana, US.

Düzen, N. E., Sahin, N., Raven, J., & Raven, C. J. (2007). The Standard progressive matrices in Turkey. In J. Raven & C. J. Raven (Eds.), *Uses and Abuses of Intelligence: Studies Advancing Spearman and Raven's Quest for Non-Arbitrary Metrics* (pp. 336–352). New York: Royal Fireworks Press.

Ekici, Ö. K. (2014). İnternet oyunları tedavi için kullanılabilir mi? *Bilim ve Teknik Dergisi*, 8.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. Retrieved from http://cdp.sagepub.com/content/11/1/19.short

Engle, R. W. (2010). Role of working-memory capacity in cognitive control. *Current Anthropology*, *51*(s1), S17–S26. doi:10.1086/650572

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In *Models of Working Memory: Models of Active Maintenance and Executive Control* (pp. 102–134). Cambridge: Cambridge University Press.

Ericsson, K. A., & Chase, W. G. (1982). Exceptional memory. *American Scientist*, *70*(6), 607–615.

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, *51*(1), 136–158. doi:10.1016/j.jml.2004.03.008

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9–15. doi:10.1111/j.1467-7687.2009.00848.x

Holmes, J., Gathercole, S. E., Place, M., Dunning, D. L., Hilton, K. A., & Elliott, J. G. (2010). Working memory deficits can be overcome : Impacts of training and medication on working memory in children with ADHD. *Applied Cognitive Psychology*, *24*, 827–836. doi:10.1002/acp

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–33. doi:10.1073/pnas.0801268105

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–49. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1546114

Kane, M. J., Bleckley, M. K., Conway, A. R. a., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169–183. doi:10.1037//0096-3445.130.2.169

Kane, M. J., & Engle, R. W. (2000). Working-Memory Capacity, Proactive Interference, and Divided Attention: Limits on Long-Term Memory Retrieval. *Journal of Experimental Psychology*, *26*(2), 336–358. doi:10.10371/0278-7393.26.2.336

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637–71. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12613671

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47–70. doi:10.1037/0096-3445.132.1.47

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*(7), 317–24. doi:10.1016/j.tics.2010.05.002

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 and Adolescent Psychiatry*, *44*(2), 177–86. doi:10.1097/00004583-200502000-00010

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, *24*(6), 781–91. doi:10.1076/jcen.24.6.781.8395

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*(1), 35–54. doi:10.1037//0033-295X.109.1.35

Mezzacappa, E., & Buckner, J. C. (2010). Working memory training for children with attention problems or hyperactivity: A school-based pilot study. *School Mental Health*, *2*(4), 202–208. doi:10.1007/s12310-010-9030-9

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, *18*(1), 46–60. doi:10.3758/s13423-010-0034-0

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience*, *7*(1), 75–9. doi:10.1038/nn1165

Raghubar, K. P., Barnes, M. a., & Hecht, S. a. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*(2), 110–122. doi:10.1016/j.lindif.2009.10.005

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., … Engle, R. W. (2012). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, 1–20. doi:10.1037/a0029082

Richmond, L. L., Morrison, A. B., Chein, J. M., & Olson, I. R. (2011). Working memory training and transfer in older adults. *Psychology and Aging*, *26*(4), 813–22. doi:10.1037/a0023631

Shah, P., & Miyake, A. (1999). Models of working memory: An introduction. In *Models of Working Memory: Models of Active Maintenance and Executive Control* (pp. 1–27). Cambridge: Cambridge University Press.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica*, *50*(3&4), 245–276.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *1-20*. doi:10.1037/a0027473

Smith, E. E., & Kosslyn, S. M. (2009). *Cognitive Psychology: Mind and Brain*. New Jersey: Pearson Education.

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The Relationship Between Working Memory and Mathematical Problem Solving in Children at Risk and Not at Risk for Serious Math Difficulties. *Journal of Educational Psychology, 96*(3), 471–491. doi:10.1037/0022-0663.96.3.471

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12*(1), 106–13. doi:10.1111/j.1467-7687.2008.00745.x

Turley-Ames, K. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, *49*(4), 446–468. doi:10.1016/S0749-596X(03)00095-0

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working wemory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *49*(1), 51–79.

Westerberg, H., Jacobaeus, H., Hirvikoski, T., Clevberger, P., Ostensson, M.-L., Bartfai, A., & Klingberg, T. (2007). Computerized working memory training after stroke - A pilot study. *Brain Injury*, *21*(1), 21–29. doi:10.1080/02699050601148726

Wykes, T., Reeder, C., Corner, J., Wuliams, C., & Everitt, B. (1999). The effects of neurocognitive remediation on executive processing in patients with schizophrenia. *Schizophrenia Bulletin*, *25*(2), 291–307.

Zhou, Q., Chen, S. H., & Main, A. (2011). Commonalities and differences in the research on children's effortful control and executive function: A call for an integrated model of self-regulation. *Child Development Perspectives*, *6*(2), no–no. doi:10.1111/j.1750-8606.2011.00176.x

APPENDICES

Appendix A. Demographic Questionnaire

1. Name:
2. Surname:
3. Gritty Games User Name:
4. Email:
5. Sex M/F
6. Birth Date:
7. Civil Status:
    - Single
    - Married
    - Separated/Divorced
8. How many children do you have, if any?
9. What is the last degree you received?
    - Did not graduate from primary school
    - Primary school
    - Middle school
    - High school
    - BA
    - MA or PhD
10. What is the last degree your mother received?
    - Did not graduate from primary school
    - Primary school
    - Middle school
    - High school

- BA
- MA or PhD

11. What is the profession of your mother?

12. What is the last degree your father received?

- Did not graduate from primary school
- Primary school
- Middle school
- High school
- BA
- MA or PhD

13. What is the profession of your father?

14. How many years of formal education did you receive?

(Do not include preschool education. You can include prep years)

15. How frequently do you play computer games?

(Please use your best guess)

- I do not play computer games.
- Less than an hour a week
- 1-3 hours a week
- 3-5 hours a week
- 5-10 hours a week
- 10-20 hours a week
- More than 20 hours a week

16. What kinds of computer games do you play? Please explain shortly.

(Game titles may help to clarify your answer)

17. What is your monthly personal expenditure?

- 0 TRY - 500 TRY
- 500 TRY – 1,000 TRY
- 1,000 TRY – 1,500 TRY
- 1,500 TRY – 2,000 TRY
- 2,000 TRY – 3,000 TRY
- 3,000 TRY – 5,000 TRY
- 5,000 TRY – 10,000 TRY
- More than 10,000 TRY

18. What is your monthly personal income?

- 0 TRY - 500 TRY
- 500 TRY – 1,000 TRY
- 1,000 TRY – 1,500 TRY
- 1,500 TRY – 2,000 TRY
- 2,000 TRY – 3,000 TRY
- 3,000 TRY – 5,000 TRY
- 5,000 TRY – 10,000 TRY
- More than 10,000 TRY

Appendix B. Forward Digit Span Numbers

3 Numbers									
1	8	1	6						
2	5	6	3						
3	2	7	8						
4 Numbers									
1	8	1	7	2					
2	3	7	2	8					
3	5	2	9	3					
5 Numbers									
1	6	8	3	4	1				
2	4	2	7	1	3				
3	2	5	1	8	7				
6 Numbers									
1	4	5	1	9	6	3			
2	3	8	2	9	1	6			
3	2	5	8	4	7	1			
7 Numbers									
1	4	5	8	6	1	7	3		
2	8	9	4	3	6	5	7		
3	8	6	4	2	9	3	2		
8 Numbers									
1	5	7	1	6	4	2	3	6	
2	3	6	7	9	2	4	8	5	
3	1	4	7	9	5	8	3	6	
9 Numbers									
1	8	9	3	6	7	2	4	1	5
2	7	4	8	6	3	2	5	9	1

3	9	3	8	7	4	6	1	2	5

Appendix C. Backward Digit Span Numbers

3 Numbers									
1	4	9	8						
2	9	5	7						
3	8	6	2						
4 Numbers									
1	2	5	6	9					
2	8	6	3	2					
3	2	5	8	1					
5 Numbers									
1	9	4	6	3	2				
2	5	9	2	4	8				
3	9	1	3	4	2				
6 Numbers									
1	8	7	5	4	6	2			
2	2	6	4	7	3	9			
3	8	3	5	2	9	6			
7 Numbers									
1	4	3	8	1	6	2	5		
2	9	1	5	8	4	7	3		
3	4	3	1	7	6	8	2		
8 Numbers									
1	4	7	8	3	6	9	1	5	
2	1	8	7	2	4	3	5	9	
3	7	4	8	5	2	6	1	3	
9 Numbers									
1	3	5	6	8	9	2	4	1	7
2	9	8	1	5	7	3	6	2	4

3	2	7	4	3	9	6	8	1	5

Appendix D. Letter Span Letters

	3 Letters								
1	B	T	E						
2	Y	N	J						
3	O	S	Z						
	4 Letters								
1	L	M	F	T					
2	R	S	N	A					
3	Y	C	M	I					
	5 Letters								
1	I	F	H	L	Z				
2	V	G	R	Y	O				
3	E	J	Y	F	D				
	6 Letters								
1	N	H	P	U	S	M			
2	U	C	P	D	J	A			
3	I	A	C	K	D	G			
	7 Letters								
1	K	N	G	Y	F	S	O		
2	M	T	Z	P	T	U	G		
3	A	P	S	R	O	C	Z		
	8 Letters								
1	E	C	B	Y	N	T	P	H	
2	T	P	L	A	H	Z	S	J	
3	I	M	J	A	O	G	N	I	
	9 Letters								
1	O	V	B	L	S	I	Y	C	G
2	P	K	G	U	T	S	L	M	R

3	F	V	E	G	K	C	H	M	U

Appendix E. Alphabet Span Letters

	3 Letters								
1	G	M	E						
2	F	D	Y						
3	U	D	M						
	4 Letters								
1	J	U	L	R					
2	N	V	C	K					
3	B	V	P	I					
	5 Letters								
1	R	O	U	F	H				
2	S	C	D	Y	M				
3	K	F	S	M	B				
	6 Letters								
1	I	R	O	L	H	N			
2	E	G	C	K	Z	F			
3	D	P	J	O	V	M			
	7 Letters								
1	C	V	U	S	O	M	P		
2	L	R	E	A	J	P	F		
3	F	E	T	J	K	Z	I		
	8 Letters								
1	U	F	R	Y	G	H	E	A	
2	Z	C	N	S	I	M	G	K	
3	O	S	B	T	K	N	I	C	
	9 Letters								
1	M	C	T	F	E	L	U	D	V
2	D	G	I	T	F	B	Y	Z	K

3	K	N	D	G	S	H	C	T	F

Appendix F. Questions in Arithmetic Test

75 seconds:

1. Portakalı ortadan kesince kaç parça olur?

2. Tuğçe'nin beş kalemi vardı. Bir kalemini kaybetti. Geriye kaç kalemi kaldı?

3. Berk'in üç lirası vardı. Babası Berk'e iki lira daha verdi. Şimdi kaç lirası var?

4. Murat'ın dokuz tane çıkartması vardı. Dayısı dört tane daha hediye etti. Murat'ın kaç çıkartması var?

5. Bir kızın on üç tane bebeği vardı, beş tanesini kardeşine verdi. Şimdi kaç bebeği kaldı?

6. Bir tane top yedi lira. Üç tane top kaç lira eder?

7. Volkan, Arif ve Kemal annelerinden dokuz lira alıyorlar. Üçünün toplam kaç parası olmuştur?

8. Bir kırtasiyede yirmi altı tane kalem vardı. Bugün on dört tane kalem satıldı. Kaç kalem kaldı?

9. Kerem bir bakkalda çalışarak otuz iki lira kazandı. Bir saatte dört lira kazandığına göre kaç saat çalışmıştır?

90 seconds:

10. Elimizde bir metrelik bir ip var. İki tane otuz beş santimetrelik parça kesiliyor. Geriye kalan ip kaç santimetredir?

11. Dört kardeşin toplam yetmiş iki lirası var. Aralarında eşit olarak paylaştıklarına göre her birinin kaç lirası vardır?

120 seconds:

12. Dört tane silgi beş lira ise yirmi dört tane silgi kaç liradır?

13. Bir kitapçı üçte bir indirim yaparak bir kitabı yirmi sekiz liraya satıyor. Kitabın indirimsiz fiyatı nedir?

14. Kırk sekiz liralık bir pantalon satılmadığı için dörtte bir indirim yapılıyor. Yine satılmayınca yarı fiyatına satılıyor. Pantalon kaç liraya satılmıştır?

15. 127 + 4 işleminin sonucu nedir?

16. 51 + 66işleminin sonucu nedir?

17. 296 + 21 işleminin sonucu nedir?

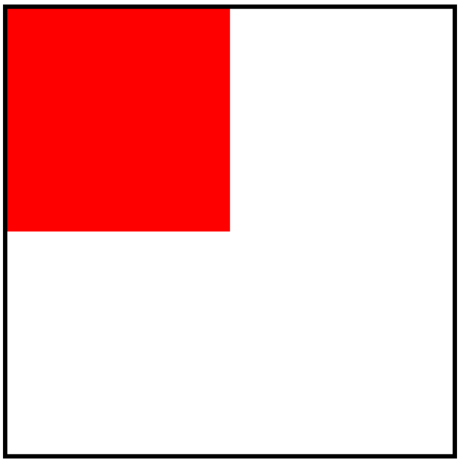
18. 628 + 931 işleminin sonucu nedir?

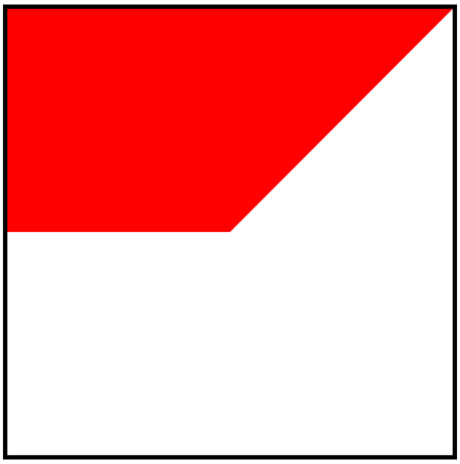
19. 2834 + 624 işleminin sonucu nedir?

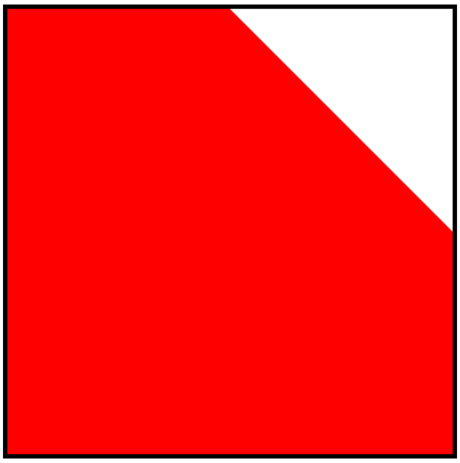
20. 1512 + 5842 işleminin sonucu nedir?

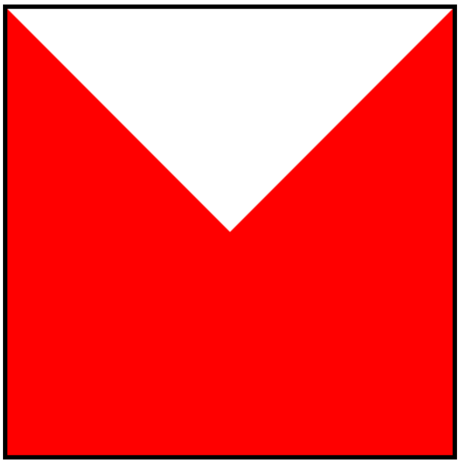
Appendix G. Images Used in the Block Design Task

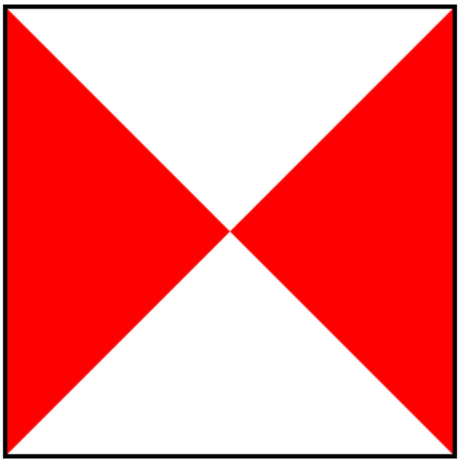
Images with four sections

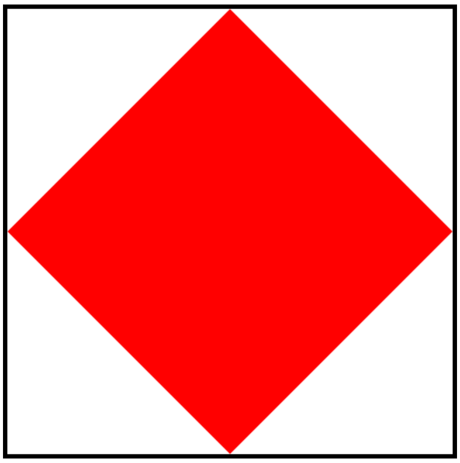


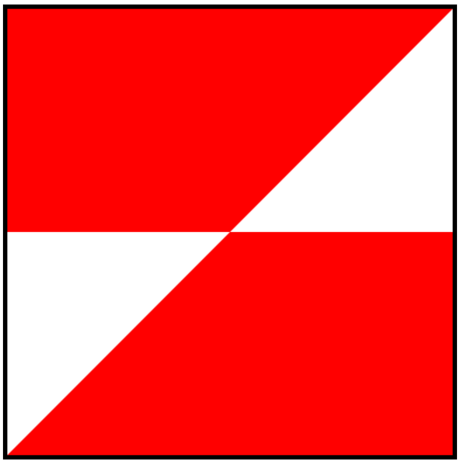


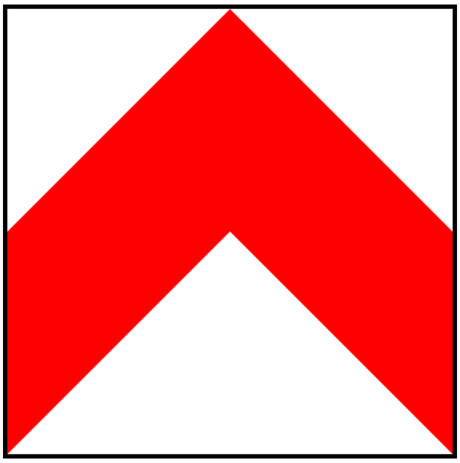




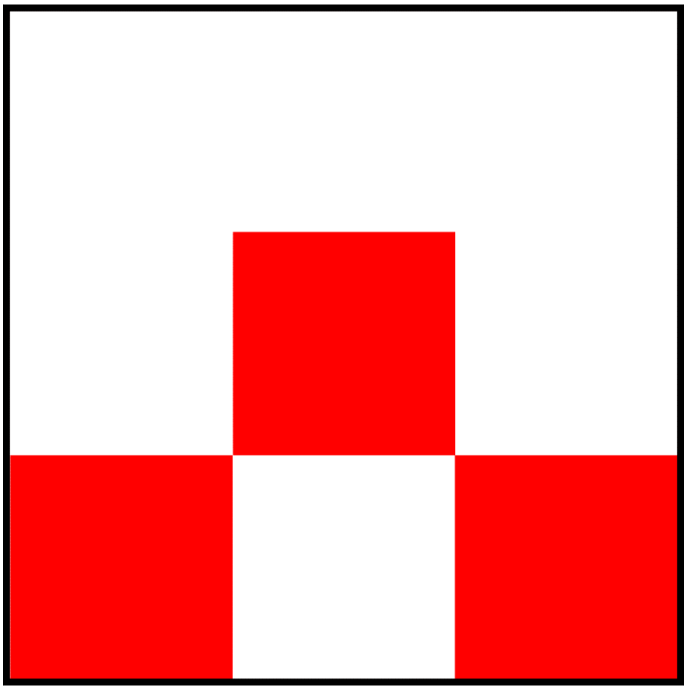


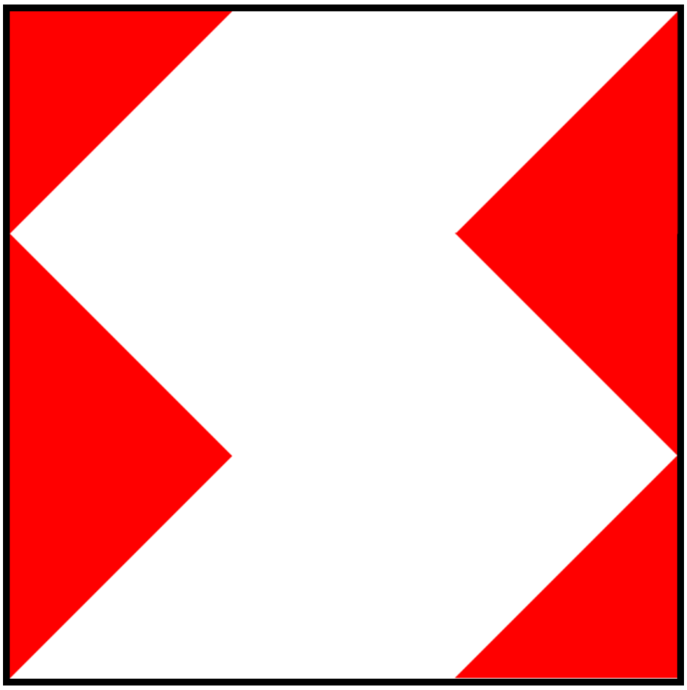


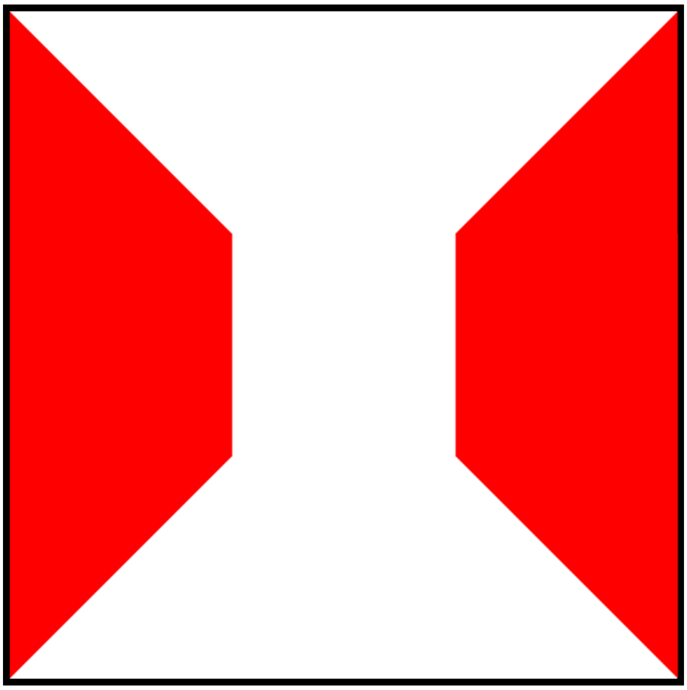


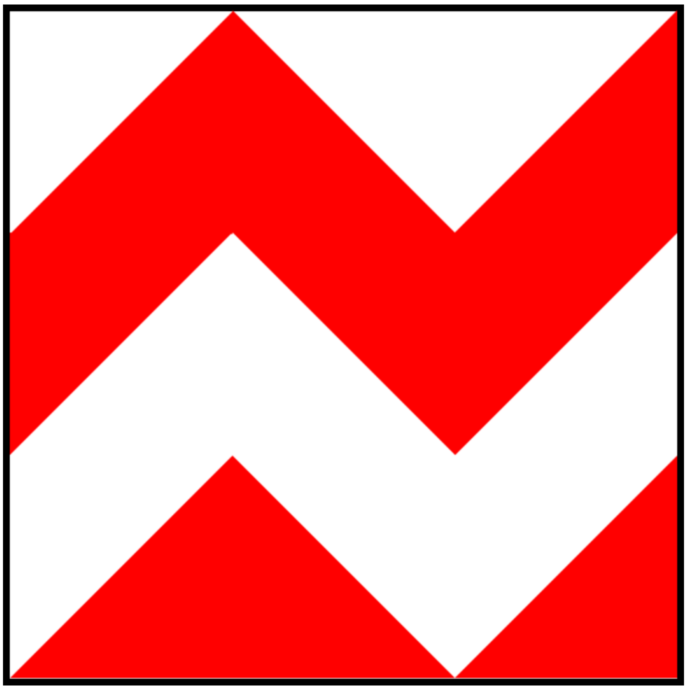


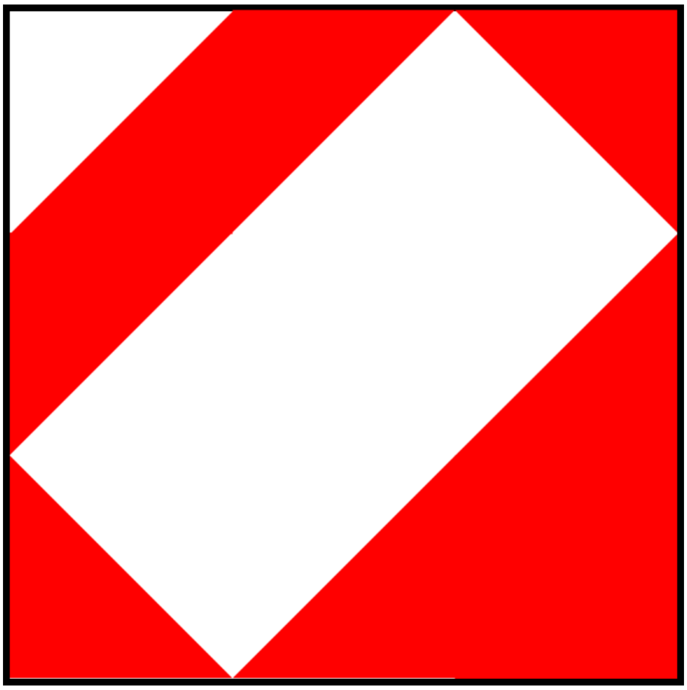
Images with Nine Sections











Appendix H. Turkish and English Names of Animals Used in the Training Task

English	Turkish
Bear	Ayı
Cat	Kedi
Chicken	Tavuk
Cow	İnek
Duck	Ördek
Fish	Balık
Monkey	Maymun
Rabbit	Tavşan
Zebra	Zebra