INVESTIGATING THE VALIDITY AND RELIABILITY OF HEMODYNAMIC MEASUREMENTS OBTAINED VIA FUNCTIONAL NEAR-INFRARED SPECTROSCOPY (*f*NIRS) BY USING N-BACK TASK

ILGIM HEPDARCAN SEZEN

MAY 2017

INVESTIGATING THE VALIDITY AND RELIABILITY OF HEMODYNAMIC MEASUREMENTS OBTAINED VIA FUNCTIONAL NEAR-INFRARED SPECTROSCOPY (*f*NIRS) BY USING N-BACK TASK

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
IZMIR UNIVERSITY OF ECONOMICS

BY

ILGIM HEPDARCAN SEZEN

MAY 2017

Approval of the Graduate School of Social Sciences

Assoc. Prof. Dr. Ö. Osman DEMİRBAŞ
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Seda DURAL
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Seda CAN
Supervisor

Examining Committee Members

Prof. Dr. Hakan ÇETİNKAYA

Assoc. Prof. Dr. Seda CAN

Assoc. Prof. Dr. Seda DURAL

ABSTRACT

INVESTIGATING THE VALIDITY AND RELIABILITY OF HEMODYNAMIC MEASUREMENTS OBTAINED VIA FUNCTIONAL NEAR-INFRARED SPECTROSCOPY (*f*NIRS) BY USING N-BACK TASK

Hepdarcan Sezen, Ilgım

Master of Science in Experimental Psychology

Supervisor: Assoc. Prof. Dr. Seda Can

May 2017

In this thesis it is aimed to investigate the validity and reliability measures and hemodynamic measures obtained through functional near infrared spectroscopy (*f*NIRS). *f*NIRS is a safe, portable, affordable, and a non-invasive device which is used in specification of the executive functions related to the hemodynamic activity of the prefrontal cortex. Verbal version of the n-back task with increasing cognitive workload from 0-, 1-, 2-, to 3-back was presented and hemodynamic measures were obtained by 16-channeled *f*NIR device. The construct validity of hemodynamic measurements was investigated based on gender-differences obtained through *f*NIRS. The test-retest reliability was examined by re-administering the same task after 3 weeks. The alternate forms reliability was investigated by using different version of the same task. Linear mixed effects model analysis indicated that for the behavioral measures, as the cognitive workload increased, accuracy decreased; but the reaction time increased. Hemodynamic measures demonstrated oxy-Hb increase mainly in left ventrolateral prefrontal cortex (VLPFC); deoxy-Hb decrease and oxygenation change in bilateral dorsolateral prefrontal cortex (DLPFC); but no significant activity of total-Hb measures was found. Construct validity examination indicated that males and females did not vary on behavioral and hemodynamic measures. Reliability

investigations by using ICCs (*intraclass coefficient*) showed consistent results both in behavioral and hemodynamic measurements. Results provide evidence for validity and reliability of behavioral measures and hemodynamic measures obtained via *f*NIRS.

Keywords: fNIR, validity, reliability, prefrontal cortex, working memory, n-back task

ÖZET

İŞLEVSEL YAKIN KIZILÖTESİ SPEKTROSKOPİDEN (*f*NIRS) ELDE EDİLEN HEMODİNAMİK ÖLÇÜMLERİN GEÇERLİK VE GÜVENİRLİĞİNİN N-GERİ GÖREVİ KULLANILARAK İNCELENMESİ

Hepdarcan Sezen, Ilgım

Deneysel Psikoloji Yüksek Lisans, Sosyal Bilimler Enstitüsü

Tez Danışmanı: Doç. Dr. Seda Can

Mayıs 2017

Bu tezde işlevsel yakın kızılötesi spektroskopiden elde edilen hemodinamik ölçümlerin geçerlik ve güvenirliğinin incelenmesi amaçlanmıştır. *f*NIRS, prefrontal kortekste gerçekleşen yürütücü işlevlerle ilişkili hemodinamik aktiviteyi belirlemede kullanılan güvenilir, taşınabilir, düşük maliyetli ve girişimsel olmayan bir cihazdır. Sözel n-geri görevi, bilişsel iş yükü 0-, 1-, 2- ve 3-geri olarak artan bir şekilde katılımcılara sunulmuştur. Hemodinamik ölçümler 16-kanallı *f*NIR cihazı aracılığıyla elde edilmiştir. Ölçümlerin yapı geçerliği cinsiyet farkı açısından davranışsal ölçümler ve *f*NIR cihazından elde edilen hemodinamik ölçümlerin ilişkisi temelinde incelenmiştir. Test-tekrar test güvenirliği aynı görevin 3 hafta sonra tekrar uygulanması, alternative form güvenirliği ise aynı görevin farklı bir formu kullanması ile incelenmiştir. Doğrusal karmaşık etkiler modeli analiz sonuçları davranışsal ölçümler için bilişsel işyükü arttıkça doğru yanıtların sayısının azaldığını; ancak tepki süresinin arttığını ortaya koymuştur. Hemodinamik ölçümler oksijenlenmiş hemoglobinin daha çok sol ventrolateral prefrontal kortekste artış gösterdiğini, oksijenden arındırılmış hemoglobin düşüşünün ve oksijenlenme değişimindeki artışın bilateral dorsolateral prefrontal kortekste meydana geldiğini; ancak toplam hemoglobinin herhangi bir anlamlı aktiviteye yol açmadığını ortaya koymuştur. Yapı geçerliği bulguları davranışsal ve hemodinamik ölçümler

bakımından kadın ve erkeklerin farklışmadığını göstermiştir. ICC (*sınıf içi korelasyon katsayısı*) aracılığıyla elde edilen güvenirlik bulguları hem davranışsal hem de *f*NIR ölçümlerinin zamanda ve farklı bir formda değişmezliğini ortaya koymuştur. Davranışsal ve *f*NIR aracılığıyla elde edilen hemodinamik ölçümlerin bulgularının benzerliği hem geçerliğe hem de güvenirliğe kanıt oluşturduğunu göstermiştir.

Anahtar Kelimeler: fNIR, geçerlik, güvenirlik, prefrontal korteks, çalışma belleği, n-geri görevi

To my only two

My sister and my dear…

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisor Assoc. Prof. Dr. Seda CAN for guiding and supporting me in a manner that is both friendly and professional, and getting involved personally when encounter with problems throughout this thesis. It was a pleasure to me to discover, explore, and learn loads of things together.

Besides my advisor, I would like to state many thanks to Prof. Dr. Hakan ÇETİNKAYA for supporting me since I was an undergraduate student, for teaching me how amazing making research and especially studying psychology is, and for giving me worthwhile advices for my academic development.

I would like to thank Assoc. Prof. Dr. Seda DURAL for sharing me insightful knowledge and comments, and for broadening my horizon since the very beginning of my master's degree studies.

I would like to acknowledge Asst. Prof. Dr. Gazihan ALANKUŞ for devotedly helping me whenever I ran into a trouble or had a question about programming my experiment.

My special thanks also goes to Assoc. Prof. Dr. Hasan AYAZ, for kindly answering my endless questions, for providing us a great opportunity to obtain *f*NIR device, and for visiting our laboratory.

I would like to thank to all my fellow lab mates Merve BULUT, Nezahat DEVECİ, Nurdan TOPRAK and Sarah HASANLİ for making the atmosphere of our laboratory as supportive and friendly as possible. Ezgi PALAZ, for settling me calmly when I got into a panic; Beste İÇAĞASI, my dear office mate, for understanding, encouraging and caring about me; Açelya YILDIZ, for believing me, for appreciating my works, for being with me, for laughing and amazing me every day. I could not forget all the good times, and cheerful memories that we have shared together. Thanks to my close friends Gamze ERDOĞAN, Aslıhan BALCI, my dear cousin Gülce HEPDARCAN, Rezzan KARLUK, and Neslihan YORTAN for their unceasing support and kind friendship for many years.

I owe a debt of gratitude to my father Ahmet HEPDARCAN, and my mother Hayat HEPDARCAN for their everlasting love, care, and support. You have

encouraged and proud of my academic researches and plans every time. My lovely sister, Gülbeniz HEPDARCAN, I would like to thank you for always being by my side, for encouraging me gladly, for criticizing me honestly, for supporting and loving me delightedly, and for and above all for being an energy source of my life.

Last but not least, Doğakan SEZEN, my dear, I would like to express my gratitude for loving, caring, and supporting me unconditionally, for encouraging me faithfully when I sink into desperation, and for watching my walk on the path of science with a pure joy and proud.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Functional near-infrared spectroscopy (*f*NIRS) is a non-invasive functional neuroimaging technique which measures hemodynamic activity changes in cerebral blood flow (CBF) in relation to the neural activity of the cortex. *f*NIRS is an optical method using near-infrared light spectroscopy by utilizing different wavelengths to obtain oxygenated hemoglobin (oxy-hemoglobin, oxy-Hb, HbO or O_2Hb) and deoxygenated hemoglobin (deoxy-hemoglobin, deoxy-Hb, HbR or Hhb) measures as an indication of neural activation in the brain (Boas, Franceschini, Dunn, & Strangman, 2002; Irani, Platek, Bunce, Ruocco, & Chute, 2007; Obrig et al., 1996; Obrig et al., 2000; Scholkmann et al., 2014; Strangman, Boas, & Sutton, 2002). Although significant advancements have been made on *f*NIRS technique, several questions depending upon validity and reliability of the measures obtained from *f*NIRS are not resolved yet. Therefore, in this thesis, the validity and reliability of hemodynamic activity measures obtained via *f*NIRS were investigated by using an n-back task.

1.1 The Historical Development of Functional Near Infrared Spectroscopy (*f*NIRS)

In objectively and precisely determining the brain-mind-behavior relationship, measurement of physiological signals of the human brain has been a crucial step. Before the progress of brain imaging techniques, the association between brain regions and cognitive functions was mostly derived from clinical neuropsychological examinations by comparing performance of healthy individuals with brain-damaged patients and post-mortem examinations (i.e., the case of Phineas Gage published by John Harlow in 1848 and 1868; the case of Mr. Leborgne ("Tan") published by Paul Broca in 1861). Latest technological advancements in brain imaging techniques bring

about significant increase in brain studies which lead researchers to conduct scientific research to explore, understand, and measure the structures, functions of the brain, neural basis of wide variety of cognitive abilities (such as attention, language and memory), and improve their understanding of brain function and/ or dysfunction in psychiatric, and neurological cases such as traumatic brain injury, Alzheimer's disease, Parkinson's disease, epilepsy, mood disorders, anxiety disorders, and personality disorders (Irani et al., 2007).

Contemporary brain-imaging techniques allow researchers to non-invasively measure structures and functions of the human brain (Zald & Curtis, 2006). Non-invasive brain imaging techniques can be categorized into two main classes in terms of whether they provide direct or indirect information about functions of the brain (Bunge & Kahn, 2009). The first class of brain imaging technique involves methods for directly measuring electrical activity related to neuronal firings, such as Electroencephalography (EEG) and Magnetoencephalography (MEG). The second main class consists of methods for indirect measurement, which functions under the principle that neural activity is supplied by hemodynamic changes resulting from electrical activity of the brain. These methods include positron emission tomography (PET), functional magnetic resonance imaging (*f*MRI), and *f*NIRS which are remarkable brain imaging techniques and they have been commonly used for studying human brain function in recent years.

As a non-invasive tool, near-infrared spectroscopy (NIRS) based optical imaging systems have been commonly used in functional brain studies in order to monitor changes in oxy-hemoglobin and deoxy-hemoglobin concentrations (Izzetoğlu, 2008). NIRS is originated by Frans Jöbsis who showed the possibility to penetrate the skull by using near-infrared light as the source (Son & Yazıcı, 2006). Studies of Jöbsis were basically about the utilization of non-invasive optical techniques to monitor intact tissues of laboratory animals. He suggested that the infrared transillumination, in the exposed heart and in the brain without surgical intervention, indicated oxygen sufficiency for cytochrome a,a_3 function, changes in tissue blood volume, and the average hemoglobin-oxy-hemoglobin equilibrium can be recorded effectively and continuously for both research and clinical purposes (Jöbsis, 1977).

The exploration of Jöbsis was followed by designing and building of several NIRS instruments. In 1980, Marco Ferrari initiated the use of a prototype of the NIRS instruments to measure changes in brain oxygenation in animal models (Ferrari et al., 1980; Giannini, Ferrari, Carpi, & Fasella, 1982) and human adults (Ferrari, Giannini, Carpi, & Fasella 1982). In 1985, Ferrari and Brazy independently demonstrated the effectiveness of first NIRS in clinical studies on the effects of carotid artery compression on regional cerebral hemoglobin oxygenation and volume cerebrovascular in newborn and adult patients (Brazy, Lewis, Mitnick, & Jöbsis, 1985; Ferrari, Giannini, Sideri, & Zanette, 1985). David Delpy, in 1984, began to develop and test several NIRS instruments. He reported the first quantitative measurement of several oxygenation and hemodynamic parameters in sick newborn infants' heads (Delpy et al, 1987) and included changes in oxy-hemoglobin, deoxy-hemoglobin and total hemoglobin (total hemoglobin = oxy-hemoglobin + deoxy-hemoglobin) concentrations, cerebral blood volume (CBV), and CBF (Wyatt, Cope, Delpy, Wray, & Reynolds, 1986; Reynolds et al., 1988). In their analyses consequently, NIRS was developed as a neuroimaging technique and named as functional NIRS (*f*NIRS) or optical topography or functional near-infrared topography (*f*NIRT) as its technology progressed (Quaresima et al., 2012).

Because *f*NIRS is an optical technique, different types of diffuse optical methods are used in the studies which are the frequency-domain technique (FD), the time-resolved systems (TRS), and continuous wave spectroscopy measurement technique (CW) (Strangman et al., 2002). Each of them is based on a specific type of illumination and, hence, has its own strengths and limitations (Ferrari, Mottola, & Quaresima, 2004; Hoshi, 2003; Izzetoğlu et al., 2004; Obrig & Villringer, 2003). The frequency-domain method illuminates the skull with intensity-modulated light. In FD method, both attenuation and phase delay of emergent light are measured to characterize the optical properties of tissue. The time-resolved system, or the time-domain technique illuminates the skull with extremely short pulses of light. The TR systems detect the shape of the pulse after diffusion through tissues and measures the temporal distributions of photons, which carry information about tissue scattering and absorption. The CW modality is based upon constant illumination of the tissue which means that it measures the light attenuation through the skull at a constant amplitude (Boas, Franceschini, Dunn, & Strangman, 2002; Izzetoğlu, 2008; Ferrari

& Quaresima, 2012). Although CW-based systems are limited to measure the amplitude attenuation of the incident light, they have several advantageous properties for certain applications: (1) CW-based measurement techniques use light emitting diodes (LEDs) rather than lasers which make them safe relating to their effects on the eyes; (2) they are low-priced compared to time and frequency domain systems (Scholkmann et al., 2014) and (3) CW-based systems can be constructed to be small, portable and wireless systems which makes them practical for research studies and clinical applications that require the quantitative measurements of hemodynamic changes during brain activation (Chance et al. 1998, Boas et al. 2002, Izzetoglu, 2008) and can be applied unobtrusively in everyday life circumstances or even freely moving animals (Muehlemann, Haensse, & Wolf, 2008).

The first *f*NIRS studies with human subjects were conducted by four research groups in 1993. They independently published their results and demonstrated that it is possible to investigate the functional brain activity non-invasively using NIRS technology (Hoshi & Tamura, 1993a, 1993b; Villringer, Planck, Hock, Schleinkofer, & Dirnagl, 1993; Chance, Zhuang, UnAh, Alter, & Lipton, 1993; Kato, Kamei, Takashima, & Ozaki, 1993).

In one of those studies, Hoshi and Tamura (1993) observed bilateral prefrontal cortex (PFC) hemoglobin changes of 14 volunteers during a mental task by using 5 single-channel CW instruments. Their study was the first study in detecting region-specific changes in both oxy-Hb and total-Hb during various mental tasks, in addition to visual and auditory stimulations. Villringer and his colleagues (1993) indicated an increased oxy-Hb and total-Hb in prefrontal cortex during a calculation task, in most (10 out 12) of the subjects similar activities were found in the occipital region during a visual stimulation task (picture watching). A study conducted by Chance and colleagues (1993) employed a different NIRS approach. They showed a cognition activated low-frequency modulation of light absorption in the PFC by utilizing a simple CW single-channel prototype. Their results revealed that the observed hemoglobin changes were related to brain activity in response to solving analogical problems that required an associative function in the frontal region. Kato and his colleagues (1993) used NIRS in order to monitor human visual cortical function throughout and after photic stimulation (PS; visual stimulation) in five adult participants. Their study was the first study which measured the adult visual cortical

areas. Results of the study showed an increase in cerebral blood volume (CBV) on the occipital surface during PS. The increase of CBV leads a rapid increase in oxy-Hb with a small increase in deoxy-Hb, suggesting cerebral vascular dilatation with decreased oxygen consumption.

In 1999, researchers from the School of Biomedical Engineering of the Drexel University (Philadelphia, PA) started to collaborate with Chance. In 2004, in order to measure adult prefrontal cortex, a wearable optical system was developed and utilized in several studies (Izzetoglu et al., 2011). In August 2009, *f*NIR Devices (Potomac, MD), using the technologies licensed from Chance and the Drexel University, started shipping a 16-channel instrument for adult PFC *f*NIRS measurements (Model 1100) (Ferrari & Quaresima, 2012). In our laboratory, *f*NIRS system that we currently utilize is a continuous wave based *f*NIR200A stand-alone functional brain imaging system which had been developed by Drexel University's Optical Engineering Team system (*f*NIR Devices, LLC (Potomac, MD, USA).

Compared to other neuroimaging techniques, *f*NIRS suggests safe, portable, low-cost methods of monitoring hemodynamic changes within the cortex. It is a compact (and some of them has a wireless) system which allows researchers to generate ecologically valid investigations about brain functions in clinical settings and also social environments without constraints related to other scanning environments, such as for elderly and multi-morbid patients whom transportation to a scanner device is hard (Hock et al., 1997; Irani et al., 2007). *f*NIRS has a real-world applicability due to its portability which enables researchers to conduct examinations in a comfortable real-world environment without intense noise or feelings of restriction. Therefore, *f*NIRS leads researchers to develop biofeedback mechanisms, which might be integrated into treatment programs (Irani et al., 2007). Furthermore, *f*NIRS has a high temporal resolution that enables researchers to obtain hemodynamic measures in 500 msec per scan. Besides its various advantages, compared to *f*MRI, *f*NIRS has a low signal-to-noise ratio (SNR) and poor spatial resolution (Hernandez-Meza et al., 2015; Gratton, Corballis, Cho, Fabiani, & Hood, 1995; Strangman, Boas, & Sutton, 2002; Totaro, Barattelli, Quaresima, Carolei, & Ferrari, 1998; Villringer & Chance, 1997; Wolf et al., 2002; Zabel & Chute, 2002).

These characteristics of *f*NIR technology make it appropriate for the study of hemodynamic changes that result in cognitive and emotional brain activity under

various research and applied settings (Izzetoğlu et al, 2004). Additionally, a variety of brain activities have been studied, which are motor activity, visual activity, auditory stimulation, olfactory, speech recognition, face recognition and the performance of cognitive tasks, by using *f*NIR technology (Obrig et al., 1996; Strangman et al., 2002; Villringer & Chance, 1997; Sakatani, Chen, Lichty, Zuo, & Wang, 1999; Bartocci, Winberg, Ruggiero, Bergqvist, Serra, & Lagercrantz, 2000; Sato, Takeuchi, & Sakai, 1999; Csibra et al., 2004).

So far, *f*NIR studies which are conducted to investigate the brain activity of humans, mainly concentrated on Brodmann's areas (BAs) 9, 10, 46, and slightly 45, 47, and 44 (Brodmann, 1909; Izzetoğlu et al., 2004) that basically corresponds to dorsolateral prefrontal cortex (DLPFC) of the human brain. DLPFC, which is one of the most important domains of prefrontal cortex, corresponds to BAs 9, 10, 45 and 46, and receives its primary inputs from the posterior parietal areas and the superior temporal sulcus (Fuchs & Philliphs, 1989; Funahashi, Bruce, & Goldman-Rakic, 1989; Kolb & Wishaw, 2009). Generally, DLPFC is associated with executive functions (D'Esposito et al., 1995; Coolidge & Wynn, 2005; Bonelli & Cummings; 2007). Executive functions comprise of a wide spectrum of high-level cognitive abilities, such as planning, initiating, sequencing (sustaining, switching, inhibiting), movement regulation in regard to environmental stimuli, changing behaviors in an appropriate way, reasoning, decision-making, complex problem-solving, long-term memory activation, motor programs construction, developing and changing strategies and the operations of working memory (Stuss & Alexander, 2000).

1.2 Underlying Physiological Principles of *f*NIR in Brain Activity Assessment

By using *f*NIRS it is intended to measure hemodynamic changes in the brain, therefore it is crucial to understand the underlying physiological principles of *f*NIR for interpretation of *f*NIRS data.

Human brain's energy reserves are small, but its energy demands are high (Hernandez-Meza et al., 2015). It consists of only 2% of total body mass. However, nearly 20% of the oxygen and 25% of the glucose are consumed by cerebral functions which demonstrate remarkable energy demands of the brain (Bélanger, Allaman, & Magistretti, 2011). The primary energy substrates that get into the brain are oxygen and glucose (Ames, 2000).

Oxygen is essential for the glucose metabolism and hence for the production of energy (Hernandez-Meza et al., 2015). Oxygen transportation occurs through the reversible bonding of oxygen to hemoglobin in the blood (Tisdall, 2007). Hemoglobin molecule is an iron-containing protein which is found in red blood cells. Each red blood cell is likely to have nearly 280 million hemoglobin molecules. The main function of hemoglobin is to transport oxygen from lungs to tissues, besides it carries carbon dioxide back to lungs (Cope, 1991). The hemoglobin molecule, as its name implies, includes *heme* which is an iron containing pigment and *globin* which is protein (Whittemore, 2004). The globin molecule of hemoglobin includes four polypeptide chains, which are α_1, α_2, β_1, and β_2 (Weissbluth, 1974). Each of the polypeptide chains binds with one heme molecule. Hence, one hemoglobin molecule contains four heme groups to which molecule of oxygen (O_2) binds (Beachey, 2012). In other words, each of the four polypeptide chains of hemoglobin molecule can bind to an oxygen molecule. Therefore, single hemoglobin can bond four oxygen molecules (Whittemore, 2004). Hemoglobin can occur in one of two forms. When hemoglobin molecule with oxygen bound to each heme it is oxy-hemoglobin. On the other hand, when hemoglobin loses any of the oxygen bound, it becomes deoxy-hemoglobin (Weissbluth, 1974; Widmaier, Raff & Strang, 2013).

Neural activity is sustained by glucose metabolism. Hence, increase in neural activity leads to increase in oxygen consumption and glucose utilization (Irani, 2007). Glucose, like oxygen and other substances in the blood, are transmitted to metabolically active neurons through blood perfusion via capillaries. A local decrease of glucose and oxygen tension stimulates the brain to dilate arterioles, which leads to an increase in CBF, and CBV. The increase in CBF and CBV results in an increase in oxy-hemoglobin and decrease in deoxy-hemoglobin in the capillary and venous compartment. The increased CBF and CBV transports both glucose and oxygen to the brain, via increased oxy-Hb in the blood which generally goes beyond the local neuronal rate of oxygen utilization, causes an overabundance of cerebral blood oxygenation in the active area (Fox, Raichle, Mintun, & Dence, 1988; Wolf et al., 2002; Irani et al., 2007; Kruggel & von Cramon, 1999). This mechanism is accepted as neurovascular coupling (Villringer et al., 1993; Obrig & Villringer, 2003). As a consequence, the increase in CBF is associated with the cerebral activity. Additionally, the very first increase in neural activity is considered as a result of a

local increase in deoxy-Hb in the capillary bed as oxygen is withdrawn from the hemoglobin (Irani, Platek, Bunce, Ruocco, & Chute, 2007).

1.3 Underlying Physical Principles of *f*NIRS in Brain Activity Assessment

Within the near-infrared light range, between 700 – 900 nm, most of the biological tissues are transparent to light (Izzetoglu et al., 2008). The near-infrared light range of wavelengths is often referred to as the 'optical window' into tissue (Boas et al., 2001; Obrig et al., 2000, Jöbsis, 1977). Within this near-infrared range, the primary light-absorbing compounds in the tissue are named as chromophores (Jöbsis, 1977). The main chromophores in the optical window are oxy-hemoglobin, deoxy-hemoglobin, water, lipids, and cytochrome-c-oxidase (Son & Yazıcı, 2006). Nearly in all near infrared spectroscopy based brain imaging studies, the main chromophores of interest are oxy-hemoglobin and deoxy-hemoglobin. This is due to orders of magnitude of other chromophores' changes are smaller than oxygenated hemoglobin and deoxygenated hemoglobin (Son & Yazıcı, 2006). Therefore, *f*NIRS utilizes specific wavelengths of light which is irradiated through the scalp to allow non-invasive measurements of oxygenated and deoxygenated hemoglobin during brain activity (Figure 1).

*f*NIRS signal is sensitive to hemodynamic changes of the cortex within the 2-3 mm and expands 1 cm laterally to either side (León-Carrión & León-Domínguez, 2012; Canning & Scheutz, 2013). Each source-detector pair follows a "banana-shaped" path (Villringer & Chance, 1997; Chance et al., 1998; León-Carrión & León-Domínguez, 2012). This path is characterized by two narrow ends and a curved inward toward the center (Ferrari & Quaresima, 2012). Photons, which pass through the cortex, go through two types of interaction: absorption and scattering. Absorption occurs due to two main chromophores of oxy-hemoglobin and deoxy-hemoglobin molecules from extracellular and intracellular boundaries of different layers of the scalp, skull, and subarachnoid space filled with cerebrospinal fluid (Izzetoglu, 2008; Ferrari & Quaresima, 2012). On the other hand, cell membranes cause scattering. When photons are absorbed by the tissue, they lose their energy. Therefore, they cannot continue to travel within the tissue.

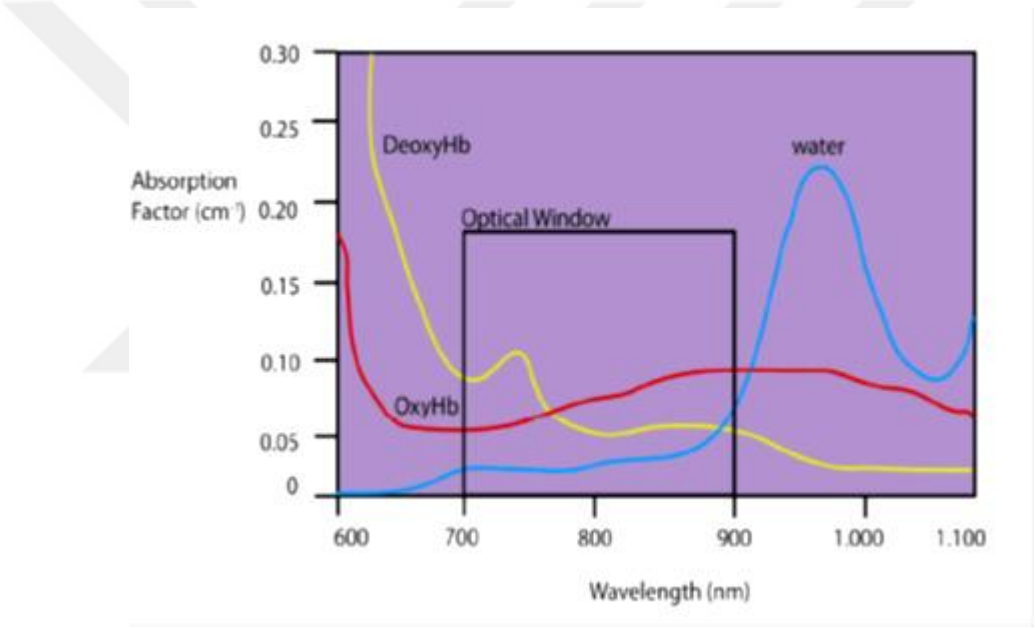

Figure 1. Absorption spectrum in optical window. In the 700 to 900 nm oxy-Hb and deoxy-Hb enables spectroscopy, whereas absorption of water becomes significant above 900 nm (Retrieved from León-Carrión, J., & León-Domínguez, U. (2012). Functional near-infrared spectroscopy (fNIRS): Principles and neuroscientific applications. In P. Bright (Ed.), Neuroimaging–Methods (pp. 47-74). Croatia: InTech).

After their entrance to the tissue, they get diffused and undergo multiple scattering events. Because the predictable quantity of photons display a "banana-shaped" profile and leave the tissue; photons can be measured by using photo-detectors. If wavelengths are selected in order to maximize the amount of absorption by oxy-hemoglobin and deoxy-hemoglobin molecules, concentration changes of these molecules cause alterations in the number of absorbed photons, besides in the number of scattered photons that leave the scalp. As a result, alterations in the concentration of these molecules give information about functions of the brain (Ayaz et al., 2010). These optical density (OD) changes at two wavelengths, oxy-hemoglobin and deoxy-hemoglobin changes, which are accessed at the surface of the scalp, are measured by using a modified Beer–Lambert Law. The original Beer–Lambert-Law is demonstrated by the equation:

$$A = \varepsilon \times (\lambda) \times c \times d$$

where, A is the absorbance of light through a medium; $\varepsilon(\lambda)$ is the absorption coefficient at wavelength λ, c, the concentration of the absorbing compound, d is the optical pathlength or distance travelled by the light beam in tissue (cm) (Obrig, & Villringer, 2003; Izzetoglu, 2008; Pellicer and Bravo, 2011). Because the Beer–Lambert Law is only valid in non-scattering media, it cannot be applied to biological tissue. In addition, the biological tissue is a highly scattering medium and does not demonstrate the true path length. Therefore, the modified Beer–Lambert Law (MBLL) was developed and commonly used in the field of near-infrared spectroscopy to take into account the light scattering (Cope & Delpy, 1988; Cope, 1991). The modified Beer-Lambert Law is expressed by the equation:

$$A = \varepsilon (\lambda) \times c \times d \times DPF (\lambda) \times + G (\lambda)$$

As in the original Beer-Lambert Law, A denotes the absorbance light attenuation; $\varepsilon (\lambda)$ is the extinction coefficient, and d is the path length of the sample. One of the new terms added to the equation is DPF which is the differential path length factor that describes the increase in path length due to tissue scattering; G is the another term that added to the equation which measures the contribution of the attenuated light due to scattering. Because the value of wavelength dependent G is usually unknown, the absolute chromophore concentration cannot be calculated.

After absorption and scattering of biological tissue, the attenuation of light intensity is denoted by the equation below:

$$I = GI_o e^{-(\alpha_{HB} C_{HB} + \alpha_{HBO2} C_{HBO2}) x L}$$

The factor that accounts for the measurement geometry is G and presumed as constant when changes occur in concentration. The input density is I_0; the molar extinction coefficients of deoxy-Hb and oxy-Hb are α_{HB} and α_{HBO2}, respectively; the concentrations of chromophores of deoxy-Hb and oxy-Hb are C_{HB} and C_{HBO2}, respectively; and L is the photon path which is a function of absorption and scattering coefficients μ_a and μ_b (Izzetoğlu., 2008).

1.4 N-back Task as a Working Memory Task and its Usage in fNIRS Studies

In most of the neuroimaging studies of working memory one of the most favored experimental paradigms is the n-back task, in which a series of single stimuli are presented and required a response when a stimulus is identical to a stimulus n position before (Owen et al., 2005; Veltman, Rombouts, & Dolan, 2003). It has been used in both experimental and neuroimaging studies as a measure of working memory ability and has been referred to as the gold-standard technique for working memory assessment in cognitive neuroscience (Owen et al., 2005, Kane & Engle, 2002; Kane, Conway, Miura, & Colflesh, 2007). In particular, N is a predetermined target and usually starts from 0-back and varied up to 3-back. In the 0-back condition, the target was pre-specified stimulus (i.e., the pre-specified letter; G) and participants were informed about that target stimulus. Thus, whenever this target stimulus was presented, participants have to respond as a target.

The 0-back condition, in some of the studies, is used to measure attentional processes of participants; because it is required from the participants to sustain their attention without working memory demand (Miller, Price, Okun, Montijo, & Bowers, 2009; Molteni et al., 2012; Lin, 2012). In other studies, 0-back condition is used as a control condition that does not require updating and manipulation of information within working memory (Braver et al., 1997; Cohen et al., 1997; Carlson et al., 1998; Nystrom et al., 2000; Rama et al., 2001; Hautzel et al., 2002; Ragland et al., 2002; Veltman, Rombouts, & Dolan, 2003; Molteni et al., 2012).

In the 1-back condition, the target stimulus was any stimulus identical to the stimulus preceding it (i.e., the target letter is presented one letter back, T-H-H, which

is H). In the 2-back condition, the target stimulus was any stimulus that was identical to the stimulus presented two trials back (i.e., the target letter is presented two trials back, N-T-N, which is N). In the 3-back condition, the target stimulus was any stimulus that was identical to the stimulus presented three trials back (i.e., the target letter is presented three trials back, G-D-N-G, which is G).

Many different types of stimuli have been utilized in n-back studies by applying various input modalities such as visual (including spatial), auditory, olfactory and vibrotactile (Carlson et al., 1998; Jansma, Ramsey, Coppola, & Kahn, 2000; Rodriguez-Jimenez et al., 2009; Jaeggi et al., 2003; Klatzky et al., 2008). Owen and his colleagues' meta-analysis study (2005) on n-back revealed that half of the n-back studies in the literature utilized verbal stimuli such as letters and words (Awh et al., 1996; Braver et al., 1997; Cohen et a., 1997; Jonides et al., 1997; Honey, Bullmore, & Sharma, 2000; Kim et al., 2002; Nystrom et al., 2000; Ragland et al., 2002; Zurowski et al., 2002, Veltman, Rombouts, & Dolan, 2003, Izzetoğlu et al., 2003; Molteni et al., 2012; Sela, Izzetoglu, Izzetoğlu, & Onaral; 2012; Herff et al., 2014), whereas the other half of the studies employed nonverbal stimuli including shapes, faces, and pictures (Carslon et al., 1998; Nystrom et al., 2000; Dade, Zatorre, Evans, & Jones-Gotman, 2001; Druzgal et al., 2001; Hautzel et al., 2002).

In either type of stimuli, n-back task requires on-line monitoring, updating, and manipulation of remembered information which places considerable demands on several key processes within working memory. Owen's meta-analysis study demonstrated that six cortical regions were consistently activated across all n-back studies (24 studies). Those regions were (1) bilateral and medial posterior parietal cortex, including precuneus and inferior parietal lobules (approximately BA7, 40); (2) bilateral premotor cortex (BA 6,8); (3) dorsal cingulate/ medial premotor cortex, including supplementary motor area (SMA; BA 32,6); (4) bilateral rostral prefrontal cortex or frontal pole (BA10); (5) bilateral dorsolateral prefrontal cortex (BA9,46); and (6) bilateral mid-ventrolateral prefrontal cortex or frontal operculum (BA45, 47). In addition, they analyzed 226 coordinates separately that were reported in 12 studies of identity monitoring of verbal stimuli and identified same six cortical regions. A separate analysis of the 76 coordinates reported by six studies of identity monitoring of nonverbal stimuli also illustrated similar activation pattern. However, this pattern was overall less salient and less global than identity monitoring of verbal stimuli.

This study revealed that n-back task can be utilized to investigate the neural basis of working memory processes due to its requirement of simultaneous monitoring of a series of stimuli, ongoing adjustment of that information to incorporate recently presented stimuli and rejection of temporally distant stimuli, and drawing of comparisons between several stimuli in the series (Owen et al., 2005).

The n-back task was created by Wayne K. Kirchner in 1958 to investigate age-related changes in reaction time and performance of young and old adults on "paced" task consisting of rapidly and continuously changing information (Kirchner, 1958). Kirchner assumed that on this type of task, participants must utilize short-term retention. He utilized visual display that involves 12 Morse keys, placed them 2 inches apart and numbered them from 1 to 12 from left to right. Above the numbers, 12 small lights were placed. Each light moved continuously from one position to another in every 1.5 seconds. Each participant was required to press a key in relation to either where the light was or had been. Four conditions were constituted: (a) no-back, in which participants were instructed to press the key where the light has appeared. The aim of this condition was to determine the ability of participants in physically keeping up with the moving light, (b) one-back, in which participants were asked to press the key where the light had just gone out, (c) two-back and (d) three-back in which participants were required to press the key where the light had gone out two positions and three positions before, respectively. The presentation order of lights was arranged in a quasi-random order fashion which was varied from trial to trial. In every trial, each sequence of lights was repeated three times. In order to warm-up the participants to the study 6-key conditions were used. However, all comparisons were made with the 12-key results. The results of the study showed that the performance of older participants was lower than younger participants. Older participants gave a less total number of responses, fewer correct responses and made more errors in 1-, 2- and 3-back conditions, compared to younger participants. Although both young and old participants were equally able to follow a physical movement of the light, older participants were unable to organize incoming and outgoing information as rapidly as the younger participants.

Braver and his colleagues (1997) studied the relationship between prefrontal cortex activity and memory load by using functional magnetic resonance imaging (*f*MRI). They investigated prefrontal cortex activity during sequential letter task by

conducting two experiments. In both experiments four conditions were used, in which memory load was increased from 0- to 3-back conditions. In their first experiment, whether any prefrontal cortex regions indicate progressively increasing activity as a function of the cognitive load was investigated. In the 0-back condition, participants were asked to respond to a single prespecified letter as a target letter (e.g., "X"). In the 1-back, 2-back and 3-back conditions, participants were required to respond any letter which was identical to the one presented one, two, and three trials back, respectively. Participants were asked to detect each letter with their dominant hand by pressing one button for targets and another for non-targets. Letters were presented 500 milliseconds (msec) from the center of the screen in a pseudo-random sequence, and the inter-stimulus-interval was arranged as 2500 msec. Behavioral results of the study were reported based on reaction time. It was found that there is a linear relationship between reaction time and cognitive load. *f*MRI results of the study also indicated a linear relationship between *f*MRI signal and cognitive load in both middle frontal gyrus (Brodmann 9/46) bilaterally and left inferior frontal gyrus (Brodmann 44/45).

At the beginning of the 2000s, Veltman and his colleagues (2003) conducted an *f*MRI study by using a maintenance task (Letter Sternberg task), and a manipulation task (n-back letter task), in order to assess both common and differential activations in the whole brain. During Sternberg task, participants were required to memorize a letter string of various lengths during 10 seconds after 15 letters presented on a screen one by one pseudo-randomly and to press one of two right-hand keys to show whether the letter had been in the string or not. A similar procedure to Braver et al., (1997) was used in n-back letter task. Results of the behavioral data of n-back task indicated that as cognitive load increased, reaction time increased and accuracy scores decreased for both tasks. Results of the brain imaging data of n-back task showed a bilateral DLPFC and ventrolateral prefrontal cortex (VLPFC), left anterior PFC, bilateral parietal cortex, left inferior temporal cortex, supplementary motor area (SMA), and cerebellum for the n-back task. The interaction of load and task type was found significant in favor of n-back task and located in bilateral DLPFC, right VLPFC, bilateral parietal cortex, SMA, and cerebellum.

In order to investigate the forebrain activation of the healthy participants during cognitive tasks, Kurtuluş İzzetoğlu and his colleagues (2003) conducted a study by using *f*NIR. Target categorization and n-back task were the cognitive tasks that were utilized in the study. In target categorization, in other words, oddball paradigm, two stimuli or classes of stimuli in a Bernoulli sequence was presented from the center of the screen. Participants were asked to count or press a button when they see the less frequent of the two events. In the n-back test, letters were chosen from the alphabet were presented on a computer screen in a random order. Four levels of difficulty were arranged. The task of the participants was to press a key when a letter repeats. If A appears, for example, followed by B, then D and D, participants must press a key when the second D appears. In the second level, participants must press a key only if a letter appears as a third in a series. If A appears, followed by B, then if another A appears participants must press a key. The inter-stimulus-interval was arranged as 2000 msec. Results of target categorization task demonstrated that oxygenation change level of the participants during infrequent targets was higher than the oxygenation change level in frequent targets. Results of the n-back task indicated that as the working memory load increased, the levels of oxygenation increased in the dorsolateral prefrontal cortex. However, it was found that in the 3-back condition oxygenation level decreased. Izzetoğlu and his colleagues made the conclusion that the demands on working memory went beyond participants' ability to follow the task. Therefore, it leads a decrease in oxygenation level when the working memory load is high.

In another *f*NIRS study, Molteni and her colleagues examined the frontal brain activity of healthy participants during a mixed attentional/ working memory task with graded levels of difficulty via time-domain *f*NIRS (Molteni et al., 2012). The aims of this study were (1) to demonstrate whether differences in frontal patterns of activation due to pure attentional and working memory tasks can be assessed by multi-channel time-domain *f*NIRS and (2) to assess time-domain *f*NIRS sensitivity in detecting modulation effects which were stimulated by the mnemonic demand of the n-back task with increasing levels of memory load. A computerized n-back task using letter version was presented to each participant. Four levels of memory load (0-, 1-, 2-, and 3-back) were used and presented twice. Therefore, 2 blocks were arranged. In the first block, participants saw the letters in 0-, 2-, 1-, and 3- back order. In the

second block, they saw the letters in different order which was 2-, 1-, 3- and 0-back. Computerized lists of 30 stimuli for each level were created and presented in a pseudo-random order. Therefore, a total of 240 intermixed letters in 24-point Helvetica form were presented to each participant in the center of the computer screen. The stimulus presentation duration was 250 msec and each stimulus was followed by a 1750 msec blank screen inter-stimulus-interval. In order to respond to a letter as a target, participants were required to press the button under their index finger (left button), to respond to a letter as a non-target, they pressed the button under their middle finger (right button). In the 0-back condition, participants were asked to respond to a single predetermined target letter (in this study, "A"). In the 1-back condition, the target letter was any letter which is identical to immediately preceding letter. In the 2- and 3-back conditions, the target letter was any letter that is identical to the letter presented two or three letters before, respectively. A pretesting session was provided to participants in order to ensure that they had understood the instructions of the study. Therefore, all participants were trained on the task. Behavioral results of the study revealed a statistically significant difference for errors made at the four different cognitive loads and for reaction time. That is, as the cognitive load increased, the number of errors and reaction time were also increased. Contrasts between 0-back and 1-, 2-, and 3-back conditions demonstrated also significant differences in activation pattern. For deoxy-Hb measures more strong and lateralized activation in the right hemisphere, compared to oxy-Hb measures was found at any load. Also, contrasts between other conditions (1-back versus 2-back, 1-back versus 3-back, etc.) indicated a trend of difference in activation patterns in the left hemisphere, although it was not significant. Researchers concluded that deoxy-Hb changes showed a strong right lateralization, whereas modifications in oxy- and total hemoglobin showed a medial localization.

In order to demonstrate whether working memory deficits results only from phonological loop (PL) or can also be seen as an outcome of poor central executive (CE) processing, Sela and her colleagues compared adult skilled and compensated dyslexic readers who have no impairment of phonological skills by using *f*NIRS (Sela, Izzetoğlu, Izzetoğlu, & Onaral; 2012). During participants' performance on n-back tasks, in which stimuli are single Hebrew consonants with conditions 0- to 3-back, their brain activity was measured. Behavioral results of their study showed no

significant group differences with respect to accuracy. Both groups' number of correct answers per block decreased as the task demand increased. A similar result was also found with regards to reaction time. That is, no significant group differences in terms of reaction time were found. Both groups' mean reaction time increased as the task demand increased. However, as opposed to behavioral results of the study, *f*NIRS results revealed significant between-group differences and task effects. Oxygenation of dyslexic readers was significantly lower than adult skilled readers, on channels 5, 7, and 8, located in the medial left frontal hemisphere. However, this oxygenation difference was observed only in 1-back condition.

In 2012, Ayaz and his colleagues revealed that *f*NIRS can be used in ecologically valid environments in order to measure mental workload levels of operators who were subjected to perform n-back and air traffic control (ATC). For n-back task, seven sessions of each of the four n-back conditions, which were 0-, 1-, 2-, and 3-back, were presented to certified professional controllers pseudo-randomly. For the ATC task, each participant was instructed to control traffic on workstations with a high-resolution radarscope, keyboard, trackball and Direct Access Keypad through 10 minutes. Two types of communications, either voice or data, were presented with a varied task difficulty by means of the number of aircraft in each sector. Participants' mental workload function on the hemodynamic response was investigated through dorsolateral and ventrolateral prefrontal cortex. Behavioral results revealed significant main effects of task difficulty for accuracy and reaction time. For accuracy, the 3-back condition was significantly lower than all n-back conditions and for reaction time, the 3-back and 2-back conditions were slower than the 1-back and 0-back conditions. *f*NIRS results indicated that as the task difficulty during task performance increase, activation level also increase, such that greater oxygenation increase in 3-back condition was found, compared to either 0-back or 1-back conditions changes only at channel 2 which corresponds to the left inferior frontal gyrus of the dorsolateral prefrontal cortex.

1.5 Psychometric Characteristics of the Measurement: Reliability and Validity of *f*NIRS Measures

As the imaging technology progresses, images of brain function are getting more complex, so today there is an increasing knowledge about processing capacities of the brain which is thought to have a correlation with the cognitive performance of

individuals. Also, more locations that the brain activity varies across trials of the experiment were obtained. The more locations were obtained the more it looks as if the whole brain involves in each cognitive process. The more it looks as if the whole brain involves in each cognitive process, the less the assumptions about functional specificity in the brain were supported. Therefore, researchers who conduct studies of brain imaging are theoretically and empirically stuck in this counter-productive circle. Consequently, they localize and modularize the functions of the brain with a bias (Hardcastle & Stewart, 2005). Nevertheless, they may be biased on conducting and publishing their brain imaging findings with this bias. Apart from localization and modularization studies, in order to measure the activity of the related brain region, many of the studies related to the brain imaging generally consider heightening the signal processing and quality of the signals (Strait & Sheutz, 2014). However, increasing the power of the devices does not mean just increasing the signal processing and quality. In order to determine whether the device has high quality, psychometric properties such as validity and reliability of the measurements obtained from the device were needed to be investigated (Scholtes, Terwee, & Poolman, 2010). In other words, whether the measurements obtained from the device measures the construct(s) that aims to measure, which refers to validity, and whether the device is free from measurement error and measurements are same for repeated measurements under various conditions, which refers to reliability must be examined. Although psychometric properties of some devices were well-documented, there is still inadequacy of information about the psychometric properties of some brain imaging devices, such as *f*NIRS (Plichta et al., 2006). The reason of insufficient information about the psychometric properties of brain imaging devices is maybe because they are seen as "engineering marvels" to observe the brain activity in the interested brain region and accepted as if they demonstrate the exact picture of psychological/and or cognitive processes. In addition, in interpreting the measurements of psychological/and or cognitive processes obtained via paper-pencil studies, psychometric properties of the related test/and or task is strictly required. Without the justification of psychometric properties of the tests, it is not possible to make conclusions about the findings of the study. So, as the validity and reliability of the scores obtained from paper-pencil tests/tasks are essential in psychological and educational assessment so the validity and reliability of the measurements obtained

via brain imaging devices are also crucial and must be assessed before using and drawing conclusions about the measurements obtained from the brain imaging device.

Therefore, studying functions of the brain using brain imaging devices by solely producing brain images is both insufficient and inaccurate. Before any measurement is conducted the validity and reliability of the measures obtained from the devices should be determined and only after the validity and the reliability have been confirmed, they should have been utilized (Franzen, 1954; Erkuş, 2003; Cook & Beckman, 2006; Scholtes, Terwee, & Poolman, 2010).

Researchers have to be sure that measurements could be replicable if the measurements were obtained from the same individuals under similar occasions, if not so, results of the studies are not trustworthy and hence the significance of the results have no meaning. The need for the consistency (stability or reproducibility) of measurements is termed as reliability. Reliability refers to the degree to which deviation scores of individuals remain relatively constant across repeated administration of the same test or alternate (parallel) test forms (Crocker & Algina, 1986). Because reliability can show variations depending on the devices/tools utilized and what is being measured, it is inevitable to investigate the reliability of the measures obtained from these devices/tools (Bennett & Miller, 2010).

The conceptual basis of reliability was derived from the classical test theory (CTT), which is a theory of measurement that determines and defines procedures toward estimating the reliability of psychological measures (Gulliksen, 1950; Magnusson, 1967). CTT reveals that the reliability of a score demonstrates the extent to which the differences in test scores of individuals are functions of their true psychological differences, as opposed to measurement error. Hence, according to CTT, reliability is originated from observed scores, true scores, and measurement error. Observed scores are referred to as values which are obtained from the measurements of some characteristics of individuals. On the contrary, true scores are referred to as values that are real amounts of that characteristic. In other words, true scores are the actual levels of the attribute that is being measured. Therefore, researchers aim to interpret observed scores of individuals as good estimates of their true score, due to their intention to demonstrate the true psychological characteristics of individuals (Furr & Bacharach, 2014). A perfect measure of a psychological characteristic means consistently assigning numbers to the characteristic that is being

measured (Murphy & Davidshofer, 2005). However, in practice, none of the measurements is perfectly reliable because measurement of any phenomenon always includes certain amount of chance error which refers to measurement error. Measurement error may result from measurement instrument itself, measurement condition, individual that carry out the measurement, or the individual being measured (Scholtes, Terwee & Poolman, 2010). Measurement error has two types which are non-random (systematic) and random (unsystematic) errors (Carmines & Zeller, 1979).

The former form of error systematically affects the measurements, which always inflates or deflates the observed score, and hence called as systematic error. Because the reliability of measurements can never be determined exactly, it can be estimated for obtained measurements (Franzen, 1954; Murphy & Davidshofer, 2005). The latter one refers to the external influences that cause inflations or deflations in scores across repeated administrations (Franzen, 1954; Furr & Bacharach, 2014). For instance, the day of the testing time can inflate some individual's score or deflate other individual's score on a test. If a high reliability coefficient (approaching 1.00) was found, that leads to consistency of measurements across different sessions, showing that the coefficient does not vary due to random errors. When errors are random, it is difficult to determine the causes of the measurement error, because they are so diverse and complicated (Murphy & Davidshofer, 2005). Therefore, the goal of reliability is to estimate errors that occur in measurement and consistently obtain stable scores from an individual in the absence of external influences (Anastasi, 1982; Franzen, 1954; Murphy & Davidshofer, 2005).

In order to estimate the reliability, several methods were used which are internal consistency, alternate (parallel) forms, test-retest method (Furr & Bacharach, 2014; Franzen, 1954). Although those methods are different methods to estimate reliability, all of them measure the amount of variance as a result of error in measurement.

Internal consistency method deals with the estimation of reliability by considering different parts of the test as different forms of that test (which is an approximation of alternate form reliability), the number of items in the test and the average intercorrelation between these items (Furr & Bacharach, 2014; Murphy & Davidshofer, 2005). Furthermore, several approaches measuring the internal

consistency of the test score can be used, which are the split-half approach, the "raw" alpha approach and the "standardized alpha" approach, which uses the basic principles of internal consistency method. If internal consistency was low, this means that the scores measure more than one construct. The consistency among the parts of the test and length of the test can affect the reliability of the measurements.

In alternate forms reliability, different forms of a test which are equivalent based on content, response processes, and statistical properties are developed (Murphy & Davidshofer, 2005). The relationship between forms is calculated and accepted as an estimate of the reliability of the measurement. If the differences between observed scores from one form and differences in the other forms are consistent, the test is accepted as reliable.

Another method of reliability estimation is the test-retest method that is widely applied in the assessment of the stability of the measurements from one test administration to the next administration (Franzen, 1954; Murphy & Davidshofer, 2005). In order to estimate reliabilities over testing sessions, intraclass correlation coefficient (ICC) is a commonly used parameter in the field of behavioral science and neuroimaging studies. The statistical basis of ICC comes from the unified analysis of variance (ANOVA) model. In other words, the appropriate ANOVA model constitutes the basis of ICC. Based on different assumptions of whether the effect of test/device is significant and whether to treat the test/device as a random or a fixed factor, different ANOVA models can be derived which are one-way random effect (model 1), two-way random effect (model 2), and two-way mixed effect model (model 3). When the effect of test is assumed to be not significant, one-way random effect was considered in which the total variance of measurements is separated into between-subjects and within-subjects variance. When the effect of the test is considered and treated as a random factor, two-way random-effect model; when the test is considered as a fixed factor both contains random factors, two-way mixed-effect model was utilized in which the total variance of measurements is separated into between-subjects, between-tests variance and random error variance in both models. For each of these three models, single measurement and average of the *k* number of measurement reliabilities can be considered. Therefore, six forms of ICCs occur, as ICC(1,1), ICC(1,k), ICC(2,1), ICC(2,k), ICC(3,1), and ICC (3,k). The first index shows one of the three ANOVA models, and the second index demonstrates

single measurement or average measurement (across *k* repeated tests) (Shrout & Fleiss, 1979). The level of significance of reliability estimates is considered as poor if ICC value is lower than 0.40, as fair if it is between 0.40 and 0.59, as good if it is between 0.60 and 0.74, and excellent if it is between 0.75 and 1.00. As a result, due to ICCs are calculated by the proportion of between-subjects variance, they should range from 0 to 1. However, due to sampling errors, ICC values can be negative. Therefore, in order to be logical, negative ICC values can be replaced by 0 (Cicchetti, 1994; Li Zeng, Lin, Cazzell & Liu, 2015). In order to choose which model should be used to estimate the test-retest reliability of measurements, properties of ICCs should be considered. If the absolute (test) agreement of the measures was interested, ICC(1,1)/ICC(1,k) and ICC(2,1)/ICC(2,k); if the consistency of measures was concerned ICC (3,1)/ICC(3,k) can be used. Nevertheless, it was recommended to choose appropriate ICCs based upon the study (Li Zeng, Lin, Cazzell & Liu, 2015). Apart from the calculation parameter of the reliability estimate, for the experimental design of the test-retest reliability, the amount of time that passes between two testing sessions has a crucial effect in test-retest method. If there is a long interval between two sessions, then there is a likelihood of greater psychological change. On the other hand, if there is a short interval between two testing sessions, carryover effects or contamination effects might occur. Therefore, it is recommended to administer the tests over a period of 2 to 8 weeks (Furr & Bacharach, 2014). The importance of each reliability estimation method varies based on the instrument that is used (Scholtes, Terwee & Poolman, 2010). For example, reliability of hemodynamic measures obtain via *f*NIRS was generally investigated by using test-retest reliability method (Watanabe, Matsuo, Kato, & Kato, 2003; Plichta et al., 2006; Plichta et al., 2012; Bhambhani, Maikala, Farag, & Rowland, 2015; Ruocco et al., 2007; Niu et al., 2013; Strangman, Goldstein, Rauch, & Stein, 2006; Wiggins, Anderson, Kitterick, & Hartley, 2016).

In one of the test-retest studies, Watanabe and his colleagues (2003) examined the test-retest reliability of bilateral frontal cortex activation of healthy participants during cognitive tasks and hyperventilation by utilizing 24-channel NIRS (12-channels in right, 12-channels left hemisphere). Researchers applied design fluency task and verbal fluency task as cognitive tasks in order to activate the frontal cortex, copy task and word repetition as control tasks to eliminate the effects of writing and

speaking, and hyperventilation as physiological stimulation in order to lead cerebral vasoconstriction. 5 of the 10 participants were retested with a retest interval of at least one year. Participants completed a copy task firstly, in which they were required to repeatedly copy the triangle shape. After that, participants were required to invent novel figures, as many as possible. Then, they were asked to repeat the letters which are told by the researcher. After this task, participants were required to generate and speak as many words as possible by beginning with a pre-specified letter. Finally, they were asked to hyperventilate as deeply as possible by following instructions of the researcher. After each of the tasks, 2 minutes resting period was arranged. Reliability of the measurements was studied by using ICC (one-way random effect model) of three different performed tasks. Results of this study showed that test-retest reliability was high in oxy-Hb measurements in verbal fluency (ICC = 0.87) and hyperventilation (ICC = 0.66), reasonable in design fluency task (0.42). However, due to the small sample size of the study (N = 5), the retest interval (205 ± 218 days) was not held constant among participants, and uncertainty of whether single or average measure ICCs were reported, results were not generalizable.

Strangman and his colleagues (2006) examined the test-retest reliability of oxy-Hb and deoxy-Hb changes only within a scanning session, over approximately 15 minutes, during a motor control task by utilizing a 32-channeled NIRS. They investigated the test-retest reliability of single blocks of a motor task, the effect of 3 different lags, which were adjacent block-pairs, for corresponding blocks across runs and delayed, between retests, and the effect of averaging on NIRS test-retest reliability. Participants were required to perform a complicated finger-to-thumb opposition task (little finger-index finger-ring finger-middle finger, then repeat). The experimental design was arranged as a block design which included 35 seconds resting period (presenting a fixation cross), then 16 seconds of periods of activity for 7 times followed by 16 seconds of resting period for 8 times, ending with a 25 second resting period. Each participant completed two runs during one scanning session, with an irrelevant motor task that occurred in between two runs. They analyzed the test-retest results by using Pearson correlation coefficients of each block. The effect of lag on test-retest correlations was modeled using a linear mixed effects model. The model was constructed as a random intercept model with lag added to model as a fixed effect. It was expected that mean correlations will differ across

participants and decline with longer lags. Findings of the study substantial block-to-block reproducibility, which lies around +0.5 based on Pearson correlation coefficients. In addition, they indicated notable consistency across oxy-Hb and deoxy-Hb measurements within a testing session. Also, they found out that averaging blocks across different lags lead significantly improved test-retest reliability. However, researchers examined the test-retest reliability only within one scanning session approximately with 15 minutes interval, in order to avoid the effects that may occur across between-sessions. Therefore, their results may be misleading due to time interval between two testing sessions. Although findings of the study revealed high variability between participants, chromophores, and comparisons between sessions at single-subject level, the test-retest reliability was found adequate at group level.

In another study, researchers compared the test-retest reliability of cerebral oxygenation and blood volume of participants who have moderate traumatic brain injury (TBI) with healthy participants throughout a rhythmic handgrip exercise with a dual wavelength NIRS (Model MRM91) (Bhambhani, Maikala, Farag, & Rowland, 2006). Each participant was tested in two separate sessions which were 24 to 48 hours apart. Before the testing session started, the maximum grip strength of right hand of the participants with a hand-grip dynamometer was measured. After their grip strength was recorded, the cerebral oxygenation and blood volume responses of participants were continuously monitored with a dual-wavelength NIRS over the left frontal lobe. The testing session was started with 2 minutes resting (baseline) state. After that, participants compressed the handgrip with their right arm maximally with one minute of intermittent, and then recover throughout 2 minutes. Participants squeezed the dynamometer maximally at a frequency of 10 compressions in every minute and they were allowed to relax along with 3 seconds in each 3 second of compression. Two-way analysis of variance was conducted in order to investigate the differences between two repeated trials. In addition, ICCs were used to calculate test-retest reliability of hand-grip strength, cerebral oxygenation, and blood volume responses of two groups. Results of this study demonstrated that ICCs for cerebral oxygenation and blood volume for healthy participants 0.83 and 0.80 and for TBI patients 0.70 and 0.64, respectively. As a result, researchers concluded that cerebral oxygenation and blood volume changes throughout hand-grip compression are

consistent in healthy and TBI patients testing times. Even though they demonstrated the reliability of NIRS during a motor task in healthy individuals and TBI patients; they did not find a statistically significant difference in cerebral blood volume between two groups. Also, the retest interval was too short as in the study of Strangman and colleagues; hence measurements may be influenced from carry-over effects.

In another study which is conducted by Plichta and his colleagues (2006) the test-retest reliability of hemodynamic measures of the occipital lobe of healthy participants, by utilizing 52-channel *f*NIRS, was examined during periodic checkerboard stimulations. Researchers focused on three different aspects to determine the reliability of the hemodynamic activity. Firstly, the variability in the hemodynamic responses was examined in single-subject level. Secondly, channel-wise comparisons were conducted. Thirdly, group comparisons were made in an activation map-wise method. In addition, whether oxy-Hb or deoxy-Hb is a more reliable parameter was investigated. In order to fulfill those aims, ICCs were used for calculating the reliability of the hemodynamic measurements across two sessions by applying one-way random-effects model. Single and average measures ICCs were reported. In addition, for the reproducibility of hemodynamic measurement changes over two sessions, channel-wise and cluster-wise (in which mean values derived from the specified channels) analyses were conducted. Moreover, quantity (with $R_{QUANTITY}$) and location (with $R_{OVERLAP}$) of the active channels for subject and group level analyses were investigated. Twelve participants took part twice in the study, with a retest interval of 3 weeks. A simple checkerboard was presented during 1200 milliseconds and reversed in contrast at 6 Hz which is followed by 13.8 seconds a black screen. Results indicated that for single measures analyses, oxy-Hb, deoxy-Hb, and total-Hb showed moderate to low reproducibility. However, it was found that single measurement reliability was improved if it is examined at a cluster level which is considered as averaging amplitudes across channels within region of interest (ROI). Especially, for oxy-Hb measurements, group level results were found highly stable across two sessions. As in single subject level, when cluster levels were considered, reliability was found sufficiently high for group level. In terms of oxy-Hb measurements, channel quantity and location were highly reproducible if a map-wise view was considered. As the researchers of the study emphasized, although they

utilized highly covering *f*NIRS (which includes 52 channels), their results were generalizable only across occipital lobe. Also, they did not utilize any specialized motion artifact removing algorithm to eliminate artefact from the data. Therefore, this may contaminate the data and lead to misevaluation of the data.

Schecklmann and his colleagues (2008) examined the short-and long-term test-retest reliability of frontotemporal brain area activity of healthy participants during a phonological version of the verbal fluency task with retest intervals of three weeks and one year using 22 channel *f*NIRS in both hemispheres. Researchers applied the verbal fluency task over three times with a retest interval of three weeks between testing session 1 and testing session 2, and a retest interval of one year between testing session 2 and testing session 3. Participants were required to produce as many nouns as possible starting with a pre-specified letter without repetitions. Fluency task was applied as a control condition, in which participants were asked to say weekdays in a consecutive manner. Each session consisted of three fluency and three control task blocks. They analyzed behavioral data by comparing a mean number of produced words with a 3 (session: session 1, session 2, session 3) x 2 (task: fluency and control task) analysis of variance (ANOVA) for repeated measures. For reproducibility, performance between sessions was correlated at the single subject level (by correlating amplitudes between sessions for each channel) and for group level (by using amplitudes within ROI). Also, single and average measure ICCs with one-way random effects model for the specific channels and for the cluster of all channels were analyzed. For the time-course reliability correlation coefficients were calculated between sessions. Lastly, quantity (with $R_{QUANTITY}$) and location (with $R_{OVERLAP}$) of the active channels were investigated. For the behavioral measurements performance was found very similar to three time points. At cluster level, acceptable reliability for all hemodynamic measure and associations across sessions were found. However, at single subject and single channel level, retest reliability was found not satisfactory. In terms of reproducibility of size and location, all correlations (correlated performances between testing sessions) were found above 0.5. In addition, they found no evidence for one specific hemodynamic measurement showed constantly higher reliability values. As a result, they concluded that group level and averaged cortical region of interest (ROI) (which was averaged across multiple channels overlying an interested cortical region) have sufficient test-retest reliability,

but added that single-subject and single-channel level analyses should be interpreted carefully. One of the reasons for finding not reproducibility of hemodynamic measure across three time points at the single-subject and single measurement level may be because of the nature of the cognitive task. Because participants were required to produce as many nouns as possible for themselves, it may create variability across participants and their response strategies and, hence, it may cause variability across brain activations. However, researchers attribute none reliability of single-subject and single-channel level findings only to differences in skull thickness and probe set localization.

In 2012, Plichta and his colleagues examined both group-level and within-subject reliability of a combined *f*MRI test battery which consists of an emotional (face matching), a motivational (monetary reward anticipation) and a cognitive (n-back working memory) task. In the face task a sequence of either fearful or angry faces (which is the experimental condition) or geometric figures (which is the control condition) were presented to participants. In each condition, each presentation includes three images (one target image (centered at the top above) and two test images (positioned left and right at the bottom)). Participants were required to identify the target image from one of the test images which is identical to the target image and have to press left or right button. In the reward task, participants had to give response based on the light-flash on the screen. The light-flash was preceded by an arrow icon with a varied direction that informs the participants about the sequence of their responses to the light-flash. Arrow up signified win condition, if the response is sufficiently fast; arrow down indicated loss condition, if the response is too slow; vertical double arrow showed verbal control, only written feedback is given, and horizontal double arrow demonstrated passive control condition in which no response is required. In the n-back task, a series of digit numbers (1-4) were presented sequentially along 500 msec with a 1500 msec inter-stimulus interval. Participants were asked to press the button if the presented number is previously highlighted number, which is the 0-back condition and two numbers back, which is the 2-back condition. 25 healthy participants were scanned in two sessions with a mean test-retest interval of 14.6 days via 3T MRI scanner. The test-retest reliability of behavioral measurements and group level consistency and within-subject reliability of *f*MRI measurements over two scanning sessions were evaluated by

using two variants of ICCs and modeled by a two-way ANOVA. In order to determine the reliability by using ICC(2,1) absolute agreement and ICC(3,1) the consistency of the measurements were calculated. Findings revealed that behavioral measurements were stable across sessions for most of the participants. For the group-level reliability of *f*MRI measurements, at the whole brain excellent consistency was found in all three tasks (ICC of reward task = 0.89; of n-back task = 0.95; of faces task = 0.98) and for within target ROIs, excellent to good consistencies were observed among scanning sessions (ICC of reward task = 0.96; of n-back task = 0.97; of faces task = 0.66). For within-subject reliability of ROI-mean amplitudes fair to good consistency for the reward task (for left ventral striatum, ICC = 0.56; for right ventral striatum, ICC = 0.62) and for the n-back task (for right DLPFC, ICC = 0.44; for middle parietal cortex, ICC = 0.57) were found, but found poor consistency for the faces task (for left amygdala, ICC = −0.02; for right amygdala ICC = 0.16) over two scanning sessions. Despite this study demonstrated task-specific reliability of the measurements, those findings were only valid for *f*MRI measurements not for *f*NIRS. Therefore, it was thought to implement the analysis method of this study in the investigation of reliability of *f*NIRS measurements.

One of the latest studies about the test-retest reliability of *f*NIRS measurements was conducted by Wiggins and his colleagues (2016). Researchers investigated the test-retest reliability of speech-evoked *f*NIRS measurements in healthy participants. Seventeen participants took part in the study twice with a three month retest interval. Three stimulation conditions were presented to participants which were auditory only, visual only (silent speech reading), and audiovisual. Participants were required to pay attention to the speech and try to understand what was said during three stimulation conditions. In order to encourage participants to sustain their attention, a control task was applied in which two words were presented on both sides of a fixation cross and required from participants to press one of the two buttons to show which of the two words had been spoken. However, measurements obtained from the control trail were not analyzed. Test-retest reliability of hemodynamic measurements was examined at group and single-subject level. Reproducibility of quantity with $R_{QUANTITY}$ and location with $R_{OVERLAP}$ of the active channels were evaluated. Reliability of measurements within ROIs was calculated using ICCs with one-way random effects model. For each ROI, both single-channel ICCs and cluster-level ICCs (mean across

channels) were considered by calculating both single and average measurements. Group level was found reliable in auditory speech condition around the pre-specified ROIs over left and right superior temporal gyri across two testing sessions. Additionally, reproducibility in both the quantity and location of significantly activate channels was found excellent for auditory-only ($R_{QUANTITY} = 1.00$, $R_{OVERLAP} = 0.83$) and audiovisual ($R_{QUANTITY} = 0.88$, $R_{OVERLAP} = 0.75$) speech conditions. However, for single-subject level, reproducibility was found highly variable. Like Schecklmann (2008) and Strangman and colleagues (2006), Wiggins and his colleagues did not found reproducibility of *f*NIRS measurements for single-subject level and offered to make advancements in optical imaging technology to improve the reliability of measurements.

As it was aforementioned, high reliability coefficient demonstrates consistency of measurements; however it does not guarantee that the inferences made from the measurements are valid. Therefore, although the first requisite to conduct a good measurement is the reliability of measurements, a reliable but an invalid measurement device does not have any importance in measurement. Hence, it is vital to investigate the validity of the measurement devices (Murphy & Davidshofer, 2004, 2005; Crocker & Algina, 1986).

In its broadest and contemporary definition, validity refers to the degree in which theory and evidence bear out the interpretations of test scores that are ensured by the proposed utilization of a test (American Education Research Association [AERA], the American Psychological Association [APA], and the National Council on Measurement in Education [NCME], 1999). From this definition, one of the implications that can be drawn is that validity is the matter of the accuracy or legitimacy of interpretations of test scores, not the test itself (Furr & Bacharach, 2014). Thus, the test itself is not either valid or invalid (Murphy & Davidshofer, 2005; Furr & Bacharach, 2014; Lissitz, 2009; Franzen, 1954). Instead of stating a test is valid or invalid, it would be said that it is valid or invalid to make inferences when it is administered to a specific individual in a specific setting because test scores themselves would be meaningless unless they refer to an observable attribute (Franzen, 1954). For instance, skin conductance of participants is accepted as a valid measurement in terms of making inferences about the conductivity of the skin of participants, such as using skin conductance to infer anxiety level of the participants

during fearful stimuli. However, skin conductance measurement will be only valid if there are no other factors, such as having a psychiatric and/ or psychological diagnosis, affect the measurement (Field, 2009). In brief, as Messick (1989) stated validity is the "scientific inquiry into test score meaning". Another, an implication of this validity definition is that validity is an issue of degree, that is, the interpretation of a test's validity should be considered as strong or weak, rather than simply stating it is valid or invalid. A measurement tool should be selected only if strong enough evidence of validity, which supports the purposed interpretation and use, is found. Theory and evidence that supports the validity of the interpretation of test scores are another important implications of this validity definition. In order to have a confidence in interpretation of measurements, there must be good empirical evidence that supports this interpretation (Furr & Bacharach, 2014).

In the field of psychological research, there is an unresolved debate about whether the validity is a tripartite or unitary concept (Franzen, 1954). From the traditional view, some of the researchers suggested that validity comprises three types which are criterion-related validity, content validity, and construct validity. Criterion-related validity refers to the correlation between tests and criteria. This type of validity was investigated in order to estimate the present standing of the individuals on the criteria (criteria refer to the interested characteristics). Differentiation of groups accordingly (which corresponds to the first type of criterion-related validity, the concurrent validity), and using test scores to predict future behavior of individuals that is assumed to show that criteria (which corresponds to the second type of criterion-related validity, the predictive validity) are associated with criterion-related validity. Content validity is about defining a universe of items and sampling those items within this universe systematically in order to establish the test. On the other hand, construct validity aims to show the attributes or traits that the test was intended to measure and to determine the degree to which specific constructs account for performance on the test (Cronbach & Meehl, 1955). From more contemporary view, construct validity is defined as a degree to which measurements obtained from the test/ instrument/device can be interpreted as it reflects the specific construct (Messick, 1987; 1989).

In order to assess construct validity, several methods can be used; which are correlating the score on the interested test and the scores on a number of other tests,

factor analysis, and experimentally manipulating the construct that is to be measured. Another method that is used in the investigation of construct validity is the multitrait-multimethod approach which is developed by Campbell and Fiske, in 1959.

Nearly in all of the validation studies of *f*NIRS, validity of the hemodynamic measurements were investigated by cross-validating NIRS measures with *f*MRI results which correspond to construct validity. The reason of comparing NIRS with *f*MRI is that both techniques are accepted as they are hemodynamic–based brain imaging techniques (i.e., Hoshi, 2003; Obrig & Villringer, 2003; Villringer & Chance, 1997) and non-invasively monitor cerebral cortical oxygenation changes (i.e., Kleinschmidt et al., 1996; Villringer et al., 1993; Hoshi & Tamura, 1993a, 1993b; Mehagnoul-Schipper et al., 2002) during hemodynamic activity. Researchers, therefore, performed studies to compare (*f*)NIRS hemoglobin measurements with BOLD *f*MRI signals in response to a variety of tasks such as motor (i.e., Kleinschmidt et al., 1996; Mehagnoul-Schipper et al., 2002; Okamoto et al., 2004; Toronov et al., 2001a; Strangman et al., 2002; Boas et al., 2003; Hoge et al., 2005; Toronov et al., 2003; Huppert, Hoge, Dale, Franceschini, & Boas, 2006a; 2006b; Huppert, Diamond, & Boas, 2008), visual tasks (Toronov, Zhang, & Webb, 2007; Zhang, Toronov, & Webb, 2006; 2005a; Abdelnour , Schmidt, & Huppert, 2009), and cognitive tasks (Cui, Bray, Bryant, Glover, & Reiss, 2010; Heinzel et al., 2013; Sato et al., 2013).

The simultaneous comparison of the MRI and NIRS methods was initially demonstrated by Kleinschmidt and his colleagues, in 1996. The aims of their study were to combine specific strengths of each brain imaging method and compare results that are acquired from simultaneous recordings of cortical activations during a sequential finger opposition task by requiring ipsi- and contralateral finger movement. Findings of their study demonstrated correlation between increase in BOLD signal of MRI and decrease in deoxy-hemoglobin concentration of NIRS in the motor cortex contralateral to the moving hand. In addition, regardless of the ipsi- and contralateral task, increase in oxy-hemoglobin concentration was similar among both MRI and NIRS measurements. However, they did not show consistent total hemoglobin changes in response to task performance. Mehagnoul-Schipper and her colleagues conducted a study in 2002, in order to compare NIRS and *f*MRI measurements in elderly participants and to identify the effect of age on cortical

oxygenation response. Researchers monitored the left motor cortex of the six healthy young and five healthy elderly participants via NIRS and blood-oxygen-level dependent (BOLD) *f*MRI simultaneously, during a contralateral finger-tapping task. Results of the study showed a good temporal relationship between NIRS and *f*MRI measurements of deoxy-Hb decrease over the left motor cortex in both young and elderly participants during a contralateral motor task performance. However, absolute baseline values of oxy-hHb, deoxy-Hb, and total-Hb were not obtained because of continuous wave NIRS technique limitations. In addition, researchers pointed out their small sample size and suggested larger sample size in order to clearly show the differences between young and elderly groups.

Apart from the motor region of the cortex and motor tasks, several different regions of interests had been determined and different tasks had been used in many of the NIRS and *f*MRI comparison studies. A study, which is conducted by Cui and his colleagues (2011), compared NIRS and *f*MRI signals across multiple cognitive tasks which are obtained over frontal and parietal regions of the cortex of the participants. While participants were scanned with *f*MRI and NIRS simultaneously, they performed four tasks which were left finger tapping, go/no-go, judgment of line orientation, and n-back working memory task using visuospatial stimuli. The study revealed significant NIRS and *f*MRI correlations, which are greater than 0, but those correlations showed wide variability, partly due to the scalp-brain distance, contrast to noise ratio (CNR), and signal level. No significant difference between any of the tasks was found. Furthermore, the spatial pattern of NIRS and *f*MRI measurements showed similarity across in four experimental tasks over the middle frontal area and the inferior parietal area. Although *f*MRI and NIRS both measure the hemodynamic activity, several factors, which are scalp-brain distance, CNR, and signal level, may affect the NIRS and *f*MRI correlations as it was shown in this study. Therefore, they could not wholly show the activation correlation in the focused region. Sato and his colleagues (2013) investigated the criterion-related validity of NIRS deoxy-Hb and *f*MRI BOLD signals during a verbal working memory and a finger tapping task over prefrontal cortex. Researchers also measured skin blood flow obtained via a laser Doppler flowmeter (LDF). The aims of the study were to show similar amplitude correlation between NIRS and *f*MRI prefrontal activity signals, and to specify the correlation between NIRS deoxy-Hb signals with the BOLD signals in the gray

matter (GM) layer and the soft tissue layer, also with the skin blood flow measured by LDF. Throughout participants' task performance, their functional activation signals were simultaneously measured by NIRS and *f*MRI. In order to generalize the results for prefrontal cortex, sensorimotor areas were also measured by administering finger tapping task. Findings of their study revealed that temporal changes in the NIRS deoxy-Hb measurements were significantly correlated with the BOLD signals in the gray matter rather than the signals in the soft tissue or LDF signals. Therefore, researchers claimed that results of their study provide supportive evidence for the validity of prefrontal cortex NIRS deoxy-Hb measurements. However, participants performed the tasks in the supine position. In spite of this, normal NIRS measurements were obtained while participants were in the sitting position. Therefore, the posture of the participants may have an effect on the results.

Even though these comparison studies of hemodynamic measurements were obtained via NIRS and BOLD signals recorded via *f*MRI indicated high correlations, there are some inconsistencies in the results of the combined NIRS and *f*MRI studies. Those inconsistencies stem from measuring different aspects of neural activity such as BOLD signal not only depends on total- deoxy-Hb of the sampled tissue, but also on venous inter-vascular oxygen saturation (Ogawa et al., 1992, 1993a; Buxton et al., 1998) and/or using instruments which have different reliabilities (Plichta et al., 2006).

As it was pointed out, many of the validation studies were conducted based upon the validity of the NIRS measures, by comparing it with *f*MRI technique, not the validity of *f*NIRS measurements itself. Although there are studies that questioned the validity of the *f*NIRS measurements, those studies are scarce in the literature. In one of the studies, the validity of the *f*NIRS measures was investigated by focusing on the interpretation of *f*NIRS measures in a situation where the relative composition of the sampled tissue is different from the standard (Merzagora et al., 2008). Researchers claimed that the value of the differential pathlength factor (DPF) which is utilized in the mBLL to take the scattering component into account may vary in a case to case. Hemodynamic activity of a male patient was measured via continuous wave *f*NIRS device (which has 4 light sources and 10 photodetectors), who suffered from a large subdural hematoma that is located laterally to the left frontal lobe. Therefore, the anterolateral portion of the left frontal lobe of the patient was ablated. Approximately, 4 years after the operation, computerized version of COWAT was

administered to the patient. The average maximum value of the oxy-Hb concentration change and the average minimum value of the deoxy-Hb concentration change were computed for each of the 16 channels. In addition to these, the temporal evolution of oxy-hemoglobin and deoxy-hemoglobin concentrations was averaged across three verbal fluency blocks (words started with the letters "F", "A", and "S") and across the channels which are sampled from intact cortical tissue. Results indicated a general increase in deoxy-Hb concentration and a general decrease in oxy-Hb concentration. The averaged hemodynamic response across the 3 verbal fluency blocks and across channels that sampled in intact cortical tissue revealed that the deoxy-Hb concentration increases and oxy-Hb concentration decrease at the beginning; some time later it has a reverse situation. However, they stated that ablated part of the patient's brain gives little information, due to cerebral cortex is absent there and consisted of a large amount of cerebrospinal fluid (CSF) compared to intact regions. In addition to this, in previous studies it is demonstrated that the presence and the thickness of the CSF layer have an important function on the distribution of light. Therefore, they concluded that different CSFs may affect the DPF coefficient, which gives an estimation of the real pathlength traveled by photons as a result of scattering, in the mBLL and hence affect the interpretations of the data.

The validity of the hemodynamic activity measures obtained via *f*NIRS was examined with the same case by Ruocco and his colleagues, in 2007. Hemodynamic activity of a patient who has a discrete evacuated subdural hematoma in the dorsolateral region of the left frontal lobe extending to insula and left temporal pole was measured by using *f*NIRS which has 4 light sources surrounded by 10 photodetectors. Three tasks were administered to the patient which are Valsalva maneouver, verbal (phonemic) fluency, and right and left-hand finger tapping over two sessions separated in time over 21 days. In the first session, the sensor pad was placed over standard EEG placements which correspond to F_7, F_{p1}, F_{p2}, and F_8. In the second session, sensor pad was positioned just anterior to positions C_3, C_Z, and C_4. Results of their study showed that during the verbal fluency task, oxy-Hb increased in the left medial area of the prefrontal cortex which corresponds to BA 9 and during the right finger tapping task, more channels are activated in the ipsilateral premotor area, which corresponds to BA 6, than lesioned contralateral cortical regions.

In another study, Ayaz and his colleagues (2014) investigated the validity of hemodynamic measures obtained from *f*NIRS by considering discriminant validity. The purpose of their study was to examine whether hemodynamic changes of the prefrontal cortex measured via *f*NIRS during neurocognitive tests is different in Amyotrophic lateral sclerosis (ALS) patients compared to healthy participants. Three neurocognitive tests were administered to both groups, which are the Number Interference Test (NIT) in which the difficulty level was increased across three trials, King-Devick Test (KDT), and a continuous performance test (CPT). After that, behavioral (accuracy and/ or reaction time) and hemodynamic measurements for both ALS and healthy group were analyzed. Behavioral measurements of NIT task and CPT task revealed no significant differences across groups. *f*NIRS measurements of NIT indicated a total hemoglobin change only on channel 12 in both groups. At the beginning of the first trial of NIT, the total hemoglobin was highest for the ALS patients; however, it decreased with subsequent trials. On the other hand, in healthy participants, total hemoglobin increased as the task trials increased. *f*NIR measures of CPT revealed that in both groups there is a significant total hemoglobin change only on channel 14; however, the total hemoglobin increased throughout the CPT task in the ALS patients compared to healthy participants. Behavioral results of the KDT task indicated that the task completion time of ALS patients is longer compared to healthy participants and *f*NIR measurements of KDT demonstrated increased total hemoglobin concentration on channels 1 and 7 for ALS patients compared to healthy participants. Overall, *f*NIRS measures present significant differences between ALS patients and healthy participants across three neurocognitive tasks and provide a measurement system to assess the cognitive decline.

As it is emphasized, the proof for the validity of the *f*NIRS measures is still inadequate. Also, the importance of the construct validity of the *f*NIRS measures is not given exactly. Even if it is examined, it is not enough; because in general, participants were selected from a patient population who had prefrontal cortex damage or cognitive impairment. Although *f*NIRS is commonly utilized in many studies conducted with Turkish samples, the validity of *f*NIRS measurements across a Turkish sample was not determined in any of the studies. Therefore, it is necessary to study the validity of the measurements obtained via *f*NIR device in this respect.

In a number of brain imaging studies, gender differences in behavioral and hemodynamic measurements were commonly examined. However, to investigate the construct validity was not the main reason in these studies. Also, results of those studies were very diverse; hence gender differences in behavioral and hemodynamic measurements have not been yet fully elucidated.

For example, Koch and her colleagues (2007) conducted an *f*MRI study in which the gender differences in the interaction of emotion and cognition was investigated. They induced negative emotion using negative olfactory stimulation and compared the performances of both male and female participants on verbal 0-back and 2-back tasks while they were scanning. Results revealed both behavioral performances of male and female participants were worse in 2-back during negative olfactory stimulation. However, *f*MRI results indicated, differences in brain activity of males and females. For females, interaction of verbal working memory and negative emotion is associated with activation in more emotion-associated areas; whereas, for males, more cognition and cognitive control areas were activated. They concluded that their results gave a new idea about a gender-specific cerebral mechanism, but they found inconsistent results of behavioral and *f*MRI measurements and could not account for it. Also, Speck and his colleagues (2000) investigated gender-specific differences in brain activation via *f*MRI during participants' performance on letter and number version of n-back task (1-back and 2-back) by presenting the same tasks two times. Behavioral data results indicated females had significantly higher accuracy and slightly slower reaction times compared to males. *f*MRI data results of the study revealed that male participants had bilateral activation mainly right hemisphere dominance, whereas female participants had left hemisphere activation predominantly.

As distinct from those studies, results of Schmidt and her colleagues (2009) studied gender differences in brain activity (using *f*MRI) and behavioral measurements during a verbal version of the n-back (0-, 1-, 2-, and 3-back) working memory task. For behavioral results, they found that accuracy measurements decreased and reaction time increased as the work load increased; for *f*MRI results they observed bilateral activation in the superior frontal gyrus (BA 6), middle frontal gyrus (BA 9 and 10), inferior frontal gyrus (BA 47) and the inferior parietal lobule (BA 40) in both genders.

In addition to these *f*MRI studies, gender-specific differences in behavioral and hemodynamic measurements were also investigated by means of *f*NIR device. Li and her colleagues (2010) aimed to show differences in prefrontal cortex activation of female and male participants during n-back task (letter version). To fulfill this purpose, they used 16-channel *f*NIRS imaging device to measure oxy-Hb, deoxy-Hb and total-Hb. Three n-back task conditions, which were 1-, 2-, and 3-back, were presented to participants for three trials randomly. Behavioral findings revealed that although accuracy and reaction time vary across cognitive loads, there was no significant effect of gender on behavioral measurements. In addition, accuracy and reaction time did not differ between genders for each cognitive load. On the other hand, gender difference was found in PFC activation under increasing cognitive demands. Oxy-Hb and total-Hb demonstrated significant gender differences under all task loads, while deoxy-Hb was not statistically significant across genders; specifically, male participants exhibited bilateral activation with slight left hemisphere dominance, whereas females showed left hemisphere activation. Furthermore, female participants indicated positive correlations between hemodynamic measurements and behavioral performance whereas; males did not exhibit significant brain–behavior correlations.

In a recently conducted study, gender differences in prefrontal cortex activity during n-back task performance of the participants was investigated by using a homemade multichannel continuous-wave NIRS instrument (Gao, Zhang, Luo, Liu & Gong, 2016). Their study consisted of six blocks and presented 0-back and 2-back conditions one by one. Behavioral results of the study indicated no gender difference in accuracy and reaction time. For the hemodynamic measurement results, during 2-back oxy-Hb measurement was found significantly higher for males compared to females. However, they added that gender-specific differences occurred at the beginning of the study. Oxy-Hb measurements of males increased at first then decreased and then came closer to oxy-Hb measurements of females in the later period. Due to the small sample size (N = 15), and unstandardized brain imaging, Gao and colleagues' findings cannot be accepted as valid. Also, similar to Koch (2007) and Li (2010), Gao and colleagues did not make an explanation about the inconsistency of behavioral and hemodynamic measurement findings.

Apart from these studies, Haut and Brach (2006) investigated gender difference in working and episodic memory which consisted of either words or faces by using *f*MRI. Their results showed that both males and females demonstrated strong and consistent evidence for lateralization based on stimulus type (words or faces) for both working and episodic memory task. When they presented words, they found greater left hemisphere activation. However, when face tasks were presented greater right hemisphere activation was observed in both males and females.

Additionally, Schmidt and her colleagues (2009) investigated the gender difference during the n-back task when participants were scanned with *f*MRI. Researchers used letter version of n-back task with four increasing cognitive loads which are 0-, 1-, 2-, and 3-back. 0-back condition was assigned as a control condition and the order of the tasks was pseudo-randomized. Behavioral results of the study indicated no gender differences between male and female participants in reaction time or accuracy of their responses. The number of correct responses decreased and the number of errors increased with increased cognitive load in all participants. *f*MRI results of the study showed no gender differences in patterns of brain activity. Bilateral activation was observed in the superior frontal gyrus (BA 6), middle frontal gyrus (BA 9 and 10), inferior frontal gyrus (BA 47) and the inferior parietal lobule in both genders.

By considering all of these, in this thesis it was aimed to investigate the validity of behavioral and hemodynamic measurements based on gender differences; the test-retest and alternate forms reliability of both behavioral and hemodynamic measures by using n-back task as a working memory task. In accordance with this purpose, the construct validity of measures was investigated based upon the relationship between behavioral measures and hemodynamic measures obtained via *f*NIRS, while male and female participants perform the n-back task. It was hypothesized as the working memory load increase, for the behavioral results, the number of detected target letters and number of detected non-target letters would decrease; but reaction time of detected target letters and reaction time of detected non-target letters would increase; for the hemodynamic measures, increase in oxy-Hb, total-Hb, and oxygenation change measures; but decrease in deoxy-Hb would occur in both genders in a similar way. Therefore, it was hypothesized that similar results

of behavioral and hemodynamic measures with respect to gender-differences would be a proof for the construct validity of measures.

The test-retest reliability was examined based upon the stability of both behavioral and hemodynamic measures by re-administering the same n-back task after 3 weeks. It was hypothesized that consistent results obtained from the investigation of the construct validity (first session) and after re-administering (second session) the same task would indicate the test-retest reliability of both behavioral and hemodynamic measures.

Furthermore, the alternate forms reliability was studied by considering the reliability of both behavioral and hemodynamic measures of the same participants, by administering the same n-back task with different letters. It was hypothesized that consistent results obtained from the investigation of the construct validity and after using a different version of the same task, would demonstrate the alternate forms reliability of both behavioral and hemodynamic measures.

CHAPTER 2

Method

2.1 Participants

A total of 50 healthy participants, 31 females and 19 males, took part in the construct validity investigation of behavioral and hemodynamic measures. The mean age of the participants was 20.40 (*SD* = 1.47) and the age range was 18-25. 47 of the participants were determined as right-handed and three of them were determined as left-handed, based on the Edinburgh Handedness Inventory results (Williams, 2013). Therefore, 3 left-handed participants were excluded from the study.

Fourteen healthy volunteers (eleven females and three males, mean age 20.64, *SD* = 1.34) who took part in the construct validity study, participated again in the test-retest and thirteen healthy volunteers (eleven female and one male, mean age 20.38, *SD* = 0.51) who took part in the construct validity study, participated again in alternate forms reliability investigations. All participants were determined as right-handed based on the Edinburgh Handedness Inventory results (Williams, 2013).

All of the participants were chosen from the Quantitative Methods in Psychology – I class, and they were informed that they will receive an extra course credit for their participation. Participation criteria of the study were determined as: (1) being a non-smoker, (2) not having currently and/or history of anemia, (3) not having history of head trauma, (4) not having currently and/or history of any neurological and psychological disorders, (5) not using currently and/or history of any psychoactive drugs, (6) not having currently and/or history related to vision, (7) not having currently and/or history of cardiovascular problems, (8) not having serious surgical operation. Participants who meet those participation criteria, participated in the study on their own, whereas participants who did not meet the participation

criteria, were required to bring someone else who meet the participation criteria. In addition, participants were required not to drink any caffeinated beverages before their participation in the study, because studies revealed that acute caffeine consumption causes constriction of cerebral vessels, reductions in cerebral blood flow, and also decrease in neurovascular coupling (Heilbronner, Hinrichs, Heinze & Zaehle, 2015; Pelligrino, Xu, & Vetri, 2010). In addition, due to blood pressure is related to dynamic changes in arterial pressure and cerebral blood flow velocity, before the study was initiated, blood pressure via tension measuring device (AEG, Model: BMG 4922); the pulse velocity and pulse oxygen saturation of participants by a finger type pulse oximeter (Contec ^{TM}, Model: CMS50D) were measured as an exclusion criteria. Blood pressure lower than or equal to 120 over 80 millimeters of mercury (mm Hg) is accepted as ideal (American Society of Hypertension, 2014). Therefore, participants whose blood pressure was higher than 120/80, did not take part in the study.

2.2 Stimuli, Apparatus, Material

2.2.1 Stimuli

6 letters were utilized as stimuli throughout the study of construct validity and test-retest reliability investigation. Those letters (D, G, K, M, T, and V) were chosen from the web page in which it generates letters randomly (http://www.dave-reed.com/Nifty/randSeq.html). As it was utilized in other studies in the literature, the font type of the letters was arranged as Helvetica (Molteni et al., 2012; Nystrom et al., 2000), and the type size was arranged as 48. For the alternate forms reliability, an alternative form of n-back task was constructed by selecting different letters (F, H, L, R, Y, and C) from the same random number generator web page.

2.2.2 Participant Information Form, Informed Consent Form and Participant Evaluation Form

Participant information form and informed consent form was prepared in order to inform participants about the purpose of the study, and the procedure that would be followed (see Appendix A and Appendix B). Informed consent form was prepared in order to inform participants about the purpose of the study, and the procedure that would be followed (see Appendix A). They were informed that they have rights to reject to participate in the study and also quit any time they want throughout the

study. In addition, it was stated that they can ask questions to the researcher about the procedure of the study.

Participant evaluation form was given to participants in order to obtain information about their current and previous neurological and psychological well-being. Participants were required to state whether they have, currently and/or, history of neurological or psychiatric disorders (e.g., Were you diagnosed with any psychological/neurological disorder?), are on medication (e.g., Are/ Were you on medication?), suffer from head trauma (e.g., Did you go through a head trauma?), have heart problems (e.g., Were you diagnosed with any heart problems?), have any surgical operation (e.g. Did you undergo a surgical operation?), have visual impairments (e.g., Do you have a severe visual impairment?). Participation history in previous experiments was also asked in order to examine their earlier knowledge about the cognitive tests, especially n-back test (e.g., Did you participate in any other experiment?). Finally, participants mentioned about their tiredness level just before the study and they rated their tiredness level on a 7 point Likert Scale (see Appendix B).

2.2.3 Apparatus and Material

Presentation of letters was designed by means of a stimulus presentation program which was written via MATLAB R2013a (The Mathworks Inc., MA, USA) on a desktop computer (TECHNO PC 750GB HDD/ 4GB RAM/ AMD FX-6100 3,3Ghz/ 1GB VGA) and presented to participants in a light and sound isolated experimental chamber via 19" LCD monitor (ACER, V193WBB). A wireless mouse was also provided to participants in order to let them to give response.

Hemodynamic activity of the prefrontal cortex of the participants was assessed by *f*NIR 200A stand-alone functional brain imaging system which was developed in the Drexel's Optical Brain Imaging Laboratory (*f*NIR Devices, LLC (Potomac, MD, USA)).

*f*NIR200A system contains *f*NIR Model 1100 which includes *f*NIR control box unit (Model 1100), AC power supply unit (attached power cable), electric cable between the *f*NIR control box and the AC power supply unit, USB cable between the *f*NIR control box main unit and the computer, DVDs with software setup files (COBI Studio software and USB drivers), 16-channel sensor pad (Model 1200 Sensor)

connected to *f*NIR control box. Complete *f*NIRS instrumentation was illustrated in Figure 2. The sensor pad is flexible and has a modular design which includes two parts: a reusable, flexible circuit board that transports necessary infrared sources and detectors, and a disposable single use cushioning material which is used to attach the sensor to the forehead of the participant. It is positioned over the forehead of the participants so that the horizontal symmetry axis (central y-axis) matches with symmetry axis of the head, in between eyes, and on the vertical axis the sensor pad is placed in compliance with the international 10-20 head marker system (Ayaz et al., 2006). The sensor pad contains light sources which is presented from the forehead of the participants via either light-emitting diodes (LEDs) or through fiber optical bundles (optode, or source), and photo-detectors (light detectors). A typical sensor pad includes four near-infrared light sources, 10 photo-detectors, and 16 channels with a source-detector separation of 2.5 cm. The penetration depth to tissue is approximately 1.25 cm, which indicates spatial resolution and accepted as low compared to other brain imaging devices (Ayaz et al., 2011; Izzetoglu, 2008). In addition, an elastic bandage is used in order to stabilize the position of the sensor pad. It is important to place the bandage tightly in order to allow the sensor to make good connection with the forehead. However, it must not be too tight, because it may constrict blood circulation. Also, a black bonnet is placed over the elastic bandage to block the penetration of ambient light into the sensor pad in order to prevent ambient light to reach the sensor and also help to fixate sensor pad's position. *f*NIR system in our laboratory was illustrated in Figure 3.

2.2.4 Stimulus Presentation Program

N-back task used in the study consisted of three trials and each trial includes four n-back conditions, which were 0-back, 1-back, 2-back and 3-back. Therefore, each of the participants was shown 12 n-back conditions in total.

In 0-back condition, participants were required to press left click if the letter that appeared on the screen was the pre-specified target letter (i.e., target letter was "D" for the first trial, "V" for the second trial, and "M" for the third trial) and press right click for the other letters, in other words non-target letters. In 1-back condition, the target letter was any letter identical to the one immediately preceding it.

Figure 2. *f*NIR system instrumentation (Retrieved from https://www.biopac.com/product/fnir-functional-near-infrared-optical-brain-imaging-system/).

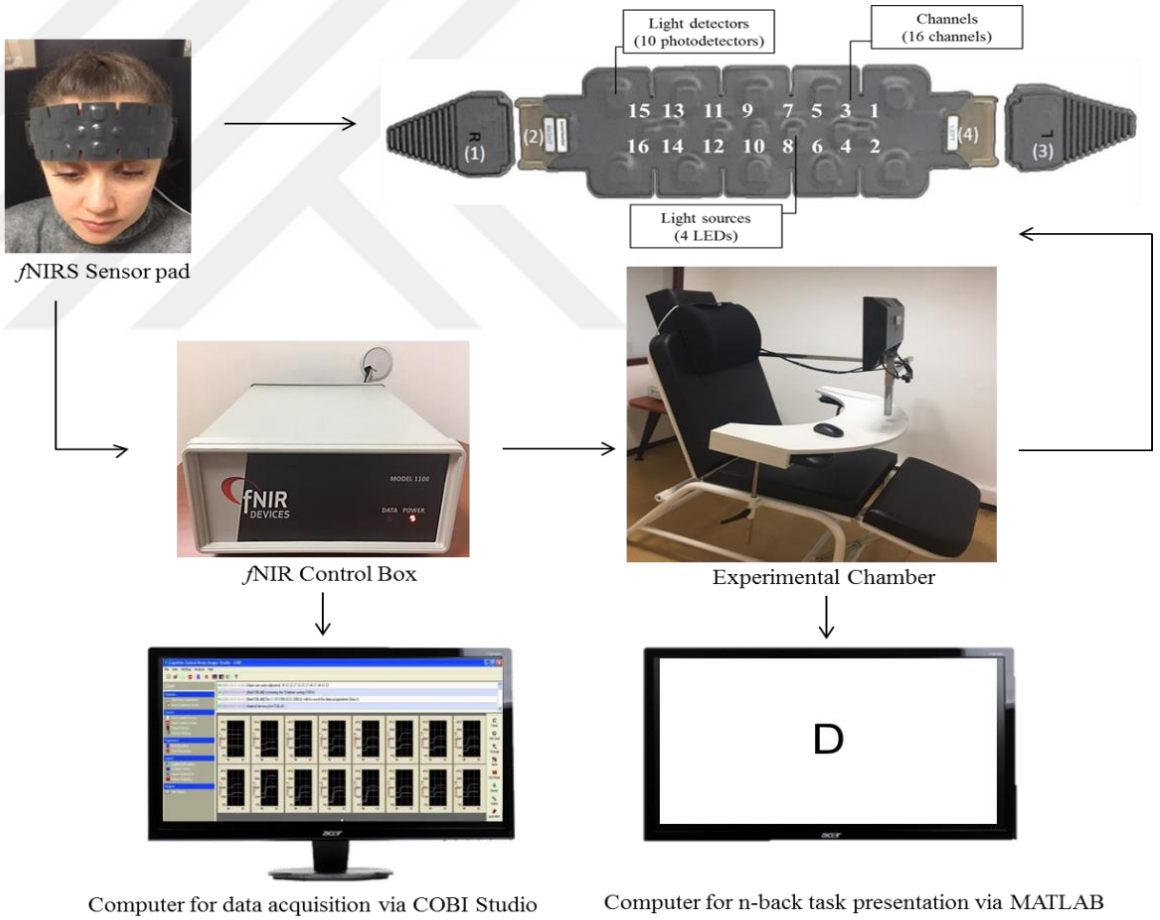

Figure 3. A complete *f*NIR system, with a sensor pad, experiment seat, control box, computer for task presentation, and COBI Studio Software included computer for data acquisition in Izmir University of Economics Neuroscience of Mind and Behavioral Research Laboratory.

Therefore, if the letter that appeared on the screen was identical to the previously appeared letter, participants were required to press left click, but if the letter which appeared on the screen was a different letter, participants were required to press right click. In 2-back condition, the target letter was any letter that was identical to the one presented two letters ago. Thus, if the letter which appeared on the screen was identical to the two letters back, participants were required to press left click, if not they were required to press right click. In 3-back condition, the target letter was any letter which was identical to the one presented three letters back. Hence, if the letter which appeared on the screen was identical to the three letters back, participants were required to press left click, otherwise participants were required to press right click. In brief, participants were required to press left click as quickly as possible if the letter that appeared on the screen was a target letter, or press right click if the letter was a non-target letter based on the instruction that was presented (Braver et al., 1997; Cohen et al., 1997; Sela, Izzetoglu, Izzetoglu, Onaral; 2012; Nystrom et al., 2000; Molteni et al., 2012; Sato et al., 2013). Figure 4 was an illustration of n-back conditions and trials for the first participant. As it was depicted in Figure 4, the order of n-back conditions was gradually increased. The target letter for the 0-back condition of the first trial is the letter "D", the letter "V" and letter "M" was determined as target letters for the second and third trial, respectively (Figure 4). Each n-back conditions included 30 letters, 10 of them were target letters, and 20 of them were non-target letters (Sela, Izzetoglu, Izzetoglu, Onaral; 2012). Hence, a total of 120 letters were presented to each of the participants.

The experimental flow was depicted in Figure 5. Prior to each condition, 1500 msec resting period was arranged. During this period, participants were required nothing but resting. By carrying out resting period, before each n-back condition, baseline light intensity levels of participants, over 15 seconds were measured via *f*NIRS (Sela, Izzetoglu, Izzetoglu, Onaral; 2012). After each resting period, prior to every single condition, a beep sound was presented on a white screen in order to take participants' attention. Beep sound was downloaded from SoundJay.com from the Sound Effects section. The duration of the sound was 1 second. After the beep sound was presented, a single letter appeared on the center of the screen for 500 msec (Braver et al., 1997; Cohen et al., 1997; Nystrom et al., 2000). After each letter disappeared, a white screen was presented for 2500 msec as an inter-stimulus-

interval (Izzetoglu et al., 2003). Then, a new letter appeared on the screen. At the end of each trial, 4000 msec inter-trail-interval was presented and then a new trial was initiated (Sela, Izzetoglu, Izzetoglu, Onaral; 2012) (Figure 5).

The stimulus presentation program was written via MATLAB R2013a and presented to participants on a 19" LCD monitor. As it was mentioned above, the stimulus presentation program consisted of three trials and each of them comprised of four n-back conditions. Presentation order of the n-back conditions was randomized based on Latin Square experimental design for the second and third trial, the orders of the n-back conditions were randomized based on the Latin Square block design. However, in the first trial, 0, 1, 2, 3-back conditions were presented to all participants respectively, since it was aimed to familiarize the participants with the n-back task by keeping the order of the first condition constant in the first trial. By applying Latin Square design, four different presentation orders were generated for the second (0-, 1-, 3-, 2-back; 1-, 2-, 0-, 3-back; 2-, 3-, 1-, 0-back; 3-, 0-, 2-, 1-back) and third trial (1-, 2-, 0-, 3-back; 2-, 3-, 1-, 0-back; 3-, 0-, 2-, 1-back; 0-, 1-, 3-, 2-back). Those constructed presentation orders were presented over in multiples of 4 (i.e., the first and fifth participant, second and sixth participant, third and seventh participant, fourth and eighth participant were shown n-back conditions in the same order) (Table 1).

2.3 Procedure

Prior to study, participants were required to read and sign Participant Information Form and Informed Consent Form (see Appendix A and Appendix B). With these forms, participants were informed about the aim of the study and the procedure that would be followed, and they would learn their rights as a participant. Then, Participant Evaluation Form was given to participants in order to gain information about whether they are appropriate to participate in the study (see Appendix C). After they fill out the Participant Evaluation Form, Edinburgh Handedness Inventory was given in order to assess the handedness of the participants (Oldfield, 1971) (see Appendix C). After participants finished filling out the forms, participants who meet the participation criteria and who accepted to participate in the study were brought to the light and sound isolated experimental chamber. Participants were placed to the experimental seat in the experimental chamber.

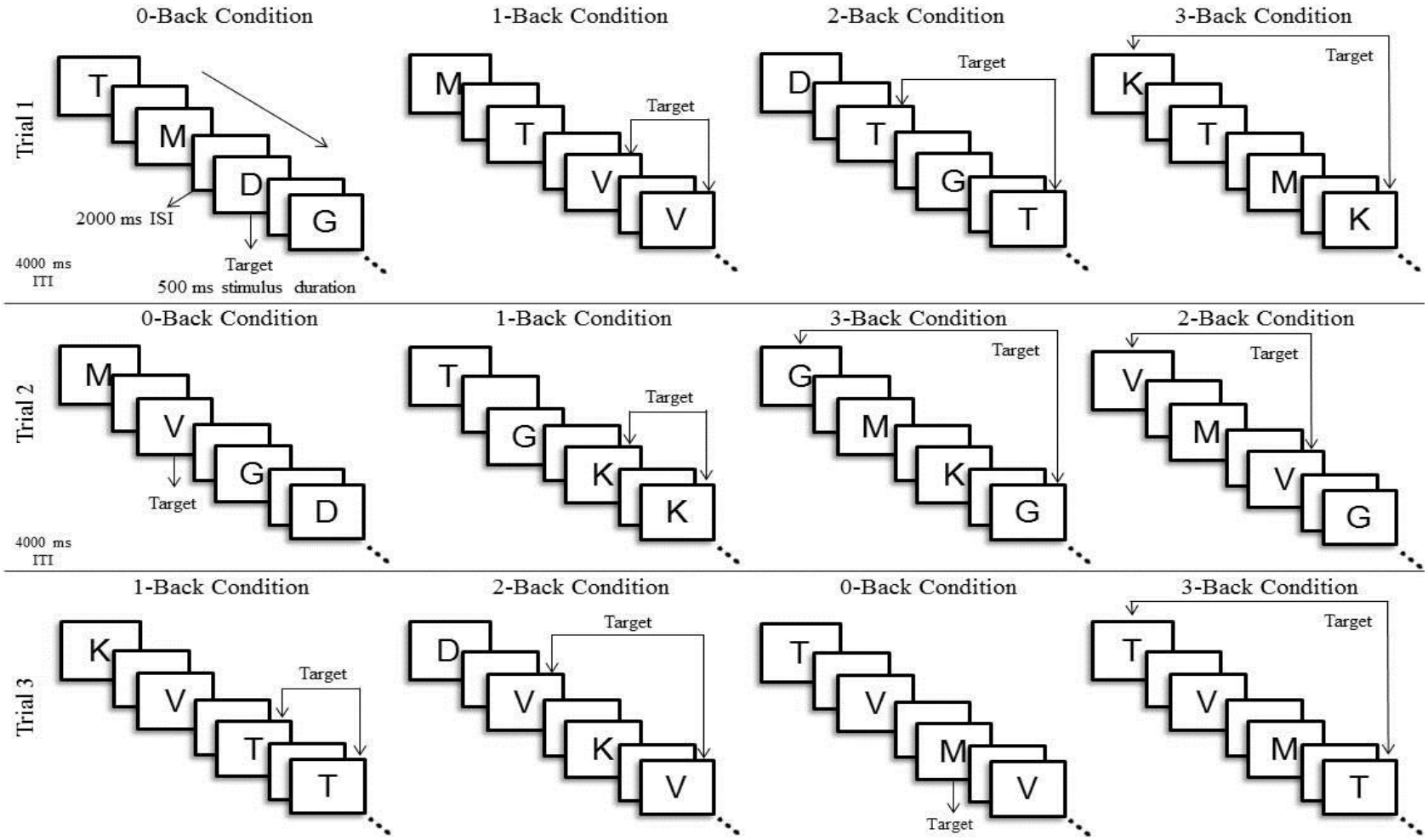

Figure 4. An illustration of all n-back conditions and trialsof the first participant.

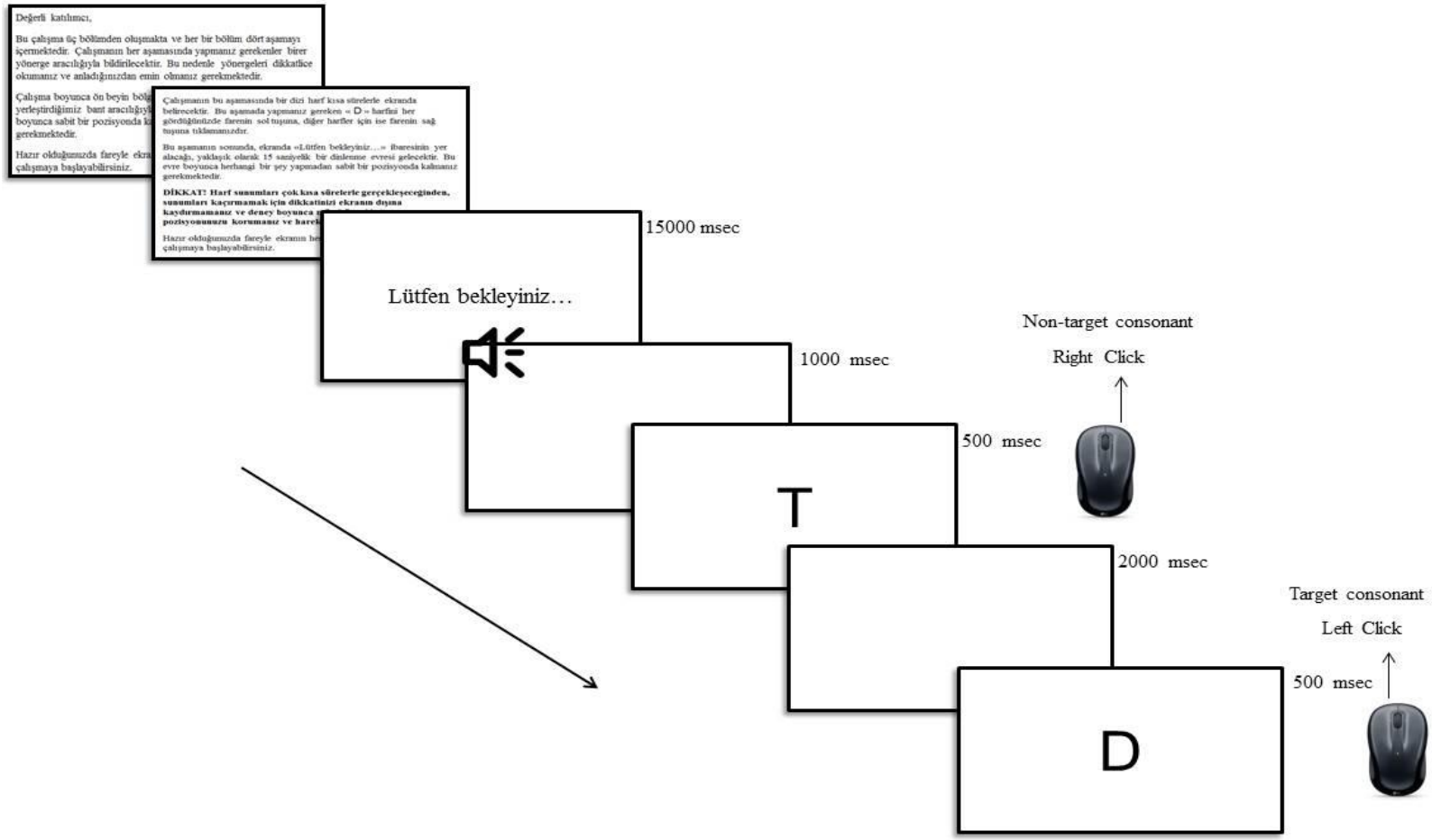

Figure 5. Experimental flow schema of the first participant during 0-back condition. The target was pre-specified as letter "D".

Table 1

Latin Square Experimental Design of the Study

		T1				T2				T3			
B1	p1	0	1	2	3	0	1	3	2	1	2	0	3
	p2	0	1	2	3	1	2	0	3	2	3	1	0
	p3	0	1	2	3	2	3	1	0	3	0	2	1
	p4	0	1	2	3	3	0	2	1	0	1	3	2
B2	p5	0	1	2	3	0	1	3	2	1	2	0	3
	p6	0	1	2	3	1	2	0	3	2	3	1	0
	p7	0	1	2	3	2	3	1	0	3	0	2	1
	p8	0	1	2	3	3	0	2	1	0	1	3	2
B3	p9	0	1	2	3	0	1	3	2	1	2	0	3
	p10	0	1	2	3	1	2	0	3	2	3	1	0
	p11	0	1	2	3	2	3	1	0	3	0	2	1
	p12	0	1	2	3	3	0	2	1	0	1	3	2
B4	p13	0	1	2	3	0	1	3	2	1	2	0	3
	p14	0	1	2	3	1	2	0	3	2	3	1	0
	p15	0	1	2	3	2	3	1	0	3	0	2	1
	p16	0	1	2	3	3	0	2	1	0	1	3	2
B5	p17	0	1	2	3	0	1	3	2	1	2	0	3
	p18	0	1	2	3	1	2	0	3	2	3	1	0
	p19	0	1	2	3	2	3	1	0	3	0	2	1
	p20	0	1	2	3	3	0	2	1	0	1	3	2
B6	p21	0	1	2	3	0	1	3	2	1	2	0	3
	p22	0	1	2	3	1	2	0	3	2	3	1	0
	p23	0	1	2	3	2	3	1	0	3	0	2	1
	p24	0	1	2	3	3	0	2	1	0	1	3	2

Table 1 (continued)

		T1				T2				T3			
B7	p25	0	1	2	3	0	1	3	2	1	2	0	3
	p26	0	1	2	3	1	2	0	3	2	3	1	0
	p27	0	1	2	3	2	3	1	0	3	0	2	1
	p28	0	1	2	3	3	0	2	1	0	1	3	2
B8	p29	0	1	2	3	0	1	3	2	1	2	0	3
	p30	0	1	2	3	1	2	0	3	2	3	1	0
	p31	0	1	2	3	2	3	1	0	3	0	2	1
	p32	0	1	2	3	3	0	2	1	0	1	3	2
B9	p33	0	1	2	3	0	1	3	2	1	2	0	3
	p34	0	1	2	3	1	2	0	3	2	3	1	0
	p35	0	1	2	3	2	3	1	0	3	0	2	1
	p36	0	1	2	3	3	0	2	1	0	1	3	2
B10	p37	0	1	2	3	0	1	3	2	1	2	0	3
	p38	0	1	2	3	1	2	0	3	2	3	1	0
	p39	0	1	2	3	2	3	1	0	3	0	2	1
	p40	0	1	2	3	3	0	2	1	0	1	3	2
B11	p41	0	1	2	3	0	1	3	2	1	2	0	3
	p42	0	1	2	3	1	2	0	3	2	3	1	0
	p43	0	1	2	3	2	3	1	0	3	0	2	1
	p44	0	1	2	3	3	0	2	1	0	1	3	2
B12	p45	0	1	2	3	0	1	3	2	1	2	0	3
	p46	0	1	2	3	1	2	0	3	2	3	1	0
	p47	0	1	2	3	2	3	1	0	3	0	2	1
	p48	0	1	2	3	3	0	2	1	0	1	3	2

Firstly, the procedure of the study was explained to participants verbally. Before the main study was initiated, a practice trial was presented to participants, in order to make them understand the task. In practice trial, only 0, 1, 2, 3-back conditions, 30 letters in each condition, were presented to each participant, respectively and three different feedbacks were provided to participants, depending on their response. If their response was correct, "You gave correct response.", if their response was incorrect, "You gave wrong response.", and if they did not give any response, "You did not give any response. Please give response for the rest of the letters either by pressing left click or right click." instructions appeared on the center of the screen. When the participants finished the practice trial, the researcher entered the experimental chamber and started to prepare the participant to the main study. This practice trial was not used in the test-retest and alternate forms applications.

The forehead of the participants was cleaned via cotton that contains make-up remover lotion, in order to remove dirt, sudation, and cosmetic materials. After the cleaning procedure was finished, then the *f*NIR sensor pad was placed on the forehead of the participants according to the international 10-20 head marker system based on nasion, inion, and the left and right preauricular points (Jasper, 1958). In order to obtain valid measurements of the *f*NIRS test-retest reliability, it was ensured to position the sensor pad as consistently as possible across participants and two testing sessions. After positioning the sensor pad to the forehead of the participants, sensor pad was immobilized via an elastic bandage and a black bonnet by properly placing to the head of the participants in order to prevent motion artifacts and also prevent the ambient light penetration (Canning & Scheutz, 2013). Participants were required to read the instructions, carefully from the computer screen and sit on the experiment seat as comfortable as possible; but preserve their positions, especially not to move their head, throughout the study. It was stated to them to feel free to ask questions before the study started, and required them to not talk until the study finished. After either side was ensured that the procedure was understood and the participant stated their readiness to start, the researcher moved to the control chamber, which was next to the experimental chamber, and opened the COBI Studio (BIOPAC Systems, Inc.) software to initiate the hemodynamic activity of the brain.

Cognitive Optical Brain Imaging (COBI) Studio is a hardware integrated software platform which allows researchers to acquire, process, and visualize *f*NIR

signals. In this thesis, COBI Studio was used in order to collect data from *f*NIR Imager hardware, via USB 2.0 ports. COBI Studio interface presents boxes which correspond to channels. The number of the channels can be increased or decreased depending on the intended purpose of the research (Figure 6). In order to start to get hemodynamic measurements, the COBI Studio has to be opened, and then new experiment has to be created by entering information about the experiment, experimenter and the participant. After that *f*NIR device can be started, however it is needed to wait for a while to get stable signals. Due to different contours, skin colors, and skull thickness of individuals, signal values may show variations. Therefore, several arrangements should be carried out for each individual before the experiment is initialized. If any of the signal values in any of the channels are high, above 4000 mV; this means that the light intensity in that channel becomes saturated. Saturation refers to the degree in which the light intensity at the detector was higher than the analog-to-digital converter limit (Ayaz, Shewokis, Izzetoğlu, Çakır, & Onaral, 2012). Higher values may be obtained because the sensor pad is not connected appropriately to the forehead of the participants and resulting in the movement of participants while the experiment was going on. Thus, the sensor pad may need to be tightened slightly in the appropriate region. However, it must not be over tight, because it may cause blood circulation constriction and hence, it may affect hemodynamic changes. Also, if any of the values in any of the channels are low, below 400 mV; this indicates that photons cannot absorb adequate light. The main reason of getting low signal values is hair under the sensor. In either case, the device has to be stopped and the values have to be arranged until appropriate signal values, which are higher than 400 mV and below 4000 mV, are obtained.

In order to get appropriate signal values, LEDs' drive current value and initial gain of the device initial may need to be changed for once for every participant in COBI Studio. It is suggested that for all channels, LEDs' drive current should have a value between 5 mA to 20 mA and the default value for initial gain of the device 20 mA. Suggested values for initial gain of the device are 1, 5, 10, 15, and 20. It is recommended that in low signal situations, when any of the values in any of the channels are below 400 mV, before increasing the gain of the device, the LEDs' drive current value should be increased.

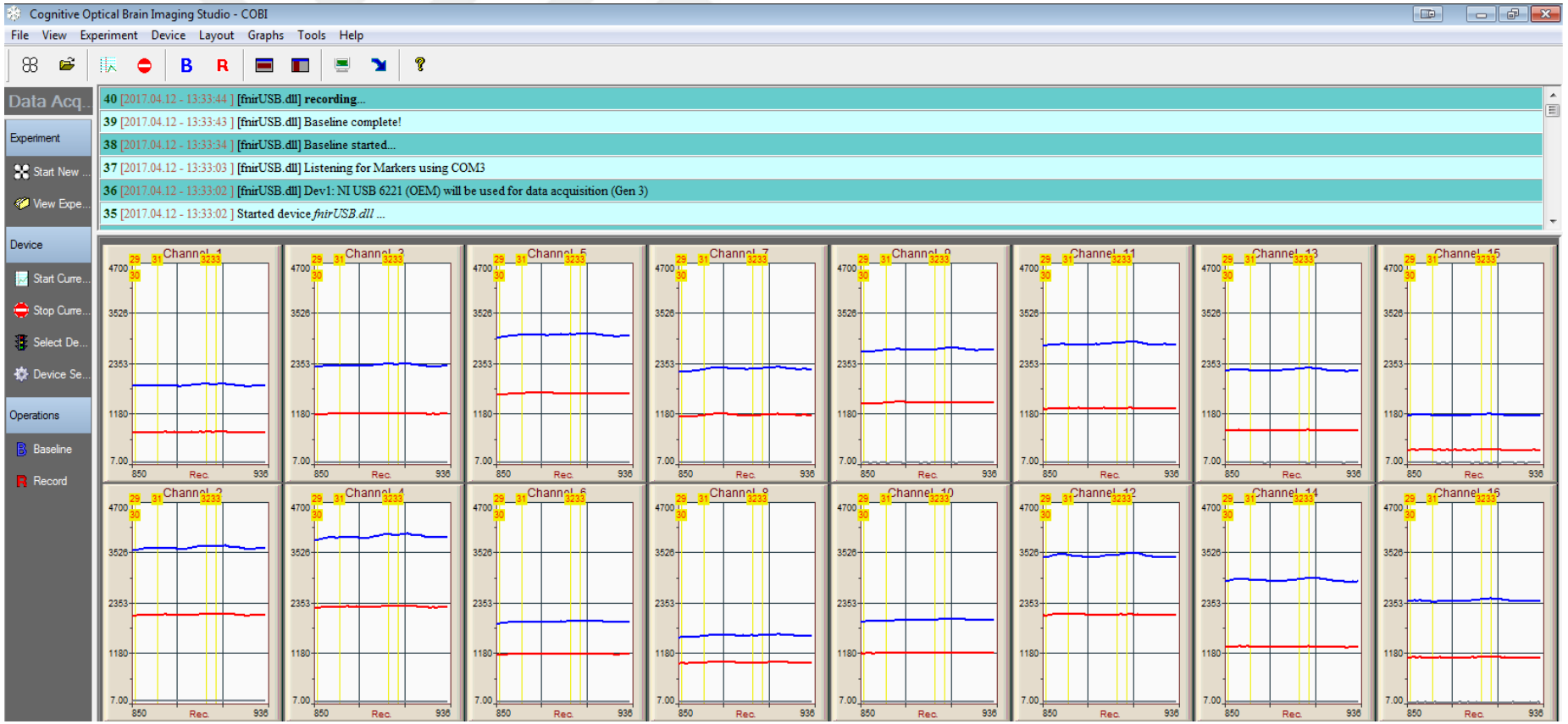

Figure 6. Illustration of hemodynamic data acquisition from COBI Studio Software. The above line (blue) in the windows, which represents channels, corresponds to 730 nm which is absorbed by deoxy-hemoglobin and the below line (red) in the windows corresponds 850 nm which is absorbed by oxy-hemoglobin. The vertical lines (yellow) correspond to time-synchronization markers.

In high signal situations, or known as saturation of the channels, in which any of the values in any of the channels are above 4000 mV, gain of the device should be decreased, before decreasing the LEDs' drive current value. After acceptable signal levels are acquired, baseline oxy-hemoglobin, deoxy-hemoglobin and raw ambient light have to be obtained, in order to be informed about the initial hemodynamic activity of the participants. Also, baseline signal values are used to calculate oxygenation via Modified Beer Lambert Law. To specify certain event times in the experiment, either manual markers can be added in COBI Studio, or pre-specified markers can be sent to COBI Studio via serial port or parallel port. In this thesis, *f*NIR data epochs for the resting and task periods were extracted from the data by using time synchronization markers via serial port. After those aforementioned device settings are arranged for each of the participant, hemodynamic activity was recorded and the stimulus presentation program from MATLAB was initiated.

After the experiment and recording is finalized, COBI Studio saves raw light intensity measurements (nir files), oxygenation values (calculated using modified Beer-Lambert Law) (oxy files) and time synchronization files (mrk files).

The first column of nir file is the time value in seconds from the beginning of the current device. The second column is the data of the first channel, in wavelength 730 nm which is absorbed mainly by deoxy-Hb. The third column is the first channel's raw ambient data. This raw ambient data refers to the dark current condition that takes the signal without any light source is on. Therefore, it enables for measurement of any possible ambient light leakage (Ayaz, Izzetoğlu, Shewokis, & Onaral, 2010). The fourth column is the data of first channel; in wavelength 850 nm that is absorbed mainly by oxy-Hb. As a result, nir file contains 3 wavelengths for all 16 channels. Therefore, in total there are 49 columns (1 time column + 3 x 16 = 49). On the other hand, the first column of the oxy file is the time value in seconds from the device has started. The rest of the columns include deoxy-Hb and oxy-Hb pairs for all 16 channels, respectively. Hence, in total there are 33 columns (1 time column + 2 x 16 = 33) (Ayaz et al, 2011, *f*NIR Devices LLC. Copyright, 2014). Besides, nir and oxy files, in the marker file the first column is the time of marker which is calculated from the start of the experiment. The second column represents the type of marker. The third column is the frame number in which the marker corresponds to. Those output files can be sent to widely-used analysis software such as MATLAB,

Excel, SPSS, SPM, *f*NIRSOFT-STD, or *f*NIRSOFT-PRO (Ayaz et al, 2011). In this thesis, the data acquired through the COBI Studio was prepared and processed for analysis by using *f*NIRSOFT-PRO analysis software.

The construct validity of behavioral (number of detected target letters, number of detected non-target letters, reaction time during target letter detection, reaction time during non-target letter detection) and hemodynamic measurements (oxy-Hb, deoxy-Hb, total-Hb, and oxygenation change measures) were investigated in terms of gender differences, while they were performing n-back task.

The test-retest reliability of both behavioral (number of detected target letters, number of detected non-target letters, reaction time during target letter detection, reaction time during non-target letter detection) and hemodynamic measures (oxy-Hb, deoxy-Hb, total-Hb, and oxygenation change) was examined by reapplying n-back task to some of the participants who took part in validity study. The retest interval was arranged as 3 weeks as Plichta and his colleagues suggested (Mean time interval = 20.36 days, *SD* = 3.14).

In order to examine the alternate (parallel) forms reliability of both behavioral (number of detected target letters, number of detected non-target letters, reaction time during target letter detection, reaction time during non-target letter detection) and hemodynamic measures (oxy-Hb, deoxy-Hb, total-Hb, and oxygenation change) n-back task was applied to participants but by utilizing different letters.

Before analyzing *f*NIR data, several steps should be followed. First of all, each of the participants' raw data should be investigated based on the quality of the signals. Quality of the signals is investigated based upon whether channels become problematic due to noises throughout the experiment. After this investigation, in order to accurately measure and analyze the hemodynamic activity in the brain, removing those determined noises from the raw *f*NIR data by applying algorithms and filters is a crucial and an essential step (Ayaz et al., 2011).

*f*NIR measures can be damaged by several physiological and/or non-physiological noise sources. In order to get more robust and reliable results, noise should be removed from the data (Izzetoglu, 2008). Noise can occur due to (1) head movements, (2) physiological signals; such as heart rate and respirations and (3) instrumental and environment-related situations (Ayaz et al., 2011). Head

movements can cause unexpected increase or decrease in the *f*NIR data; therefore it is easy to recognize this type of noise, due to its sudden variations and large spikes. Low values can be recorded due to no light pass towards the head or extremely high values can be obtained because of the reflection of light from the surface of the head (Ayaz, Izzetoğlu, Shewokis, & Onaral, 2010). If any of the signal values in any of the channels are above the 4000 mV, this signifies high values and then it means that the channel is saturated. Although saturation can be detected when the COBI Studio is initialized, it can also occur during the flow of the hemodynamic activity, but because of the ongoing activity, metabolic alterations in the blood cannot be manipulated. Therefore, motion artifacts from the raw *f*NIR data have to be eliminated after the experiment is finalized. Physiological signals, like heart rate (above 0.5 Hz) and respiration (above 0.2 Hz) are other types of noises and have higher frequencies compared to hemodynamic responses. Therefore, they generate signal drifts in the *f*NIR data. Another type of noise, which are instrument and environmental noises, can result from ambient light; such as daylight, room light, and also light from computer monitor. Suggested way to eliminate this type of noise is to prepare the experimental environment as DC direct current lighting, room light as 60 Hz and computer monitor as 60-75 Hz (Ayaz, 2011). In addition, the temperature of the experimental chamber may affect the signal quality; therefore it is recommended to set experimental chamber temperature to 16^{o}C as it was articulated by Hasan Ayaz. Before the hemodynamic data obtained via COBI Studio was analyzed, raw *f*NIR data (nir formatted data file) with the related marker (event) files were loaded to *f*NIRSOFT-PRO analysis software (Ayaz et al., 2010). In *f*NIRSOFT-PRO, researchers also have an option to load oxy formatted data, in which the raw light data had been already converted to oxy-Hb and deoxy-Hb with modified Beer-Lambert Law which is obtained through COBI Studio. In this thesis, nir formatted data files were loaded, prepared and processed; because starting from loading nir files to *f*NIRSOFT-PRO offers different algorithms to be applied, that helps to remove various artifacts from the raw data and to improve signal quality.

Automation of poor-signaled channels, improvement of signals and reduction of dynamic noises lead to development of various algorithms and filters in NIRS based studies which constitutes the first step for processing *f*NIRS data (Cui, Bray, & Reiss, 2010; Harrivel & Hearn, 2012; Ayaz, 2011).

Artifact removal approaches can be classified into three categories (Cui, Bray, & Reiss 2010). Removing artifacts based on its temporal characteristics is the first category and includes methods such as low-pass filtering, high-pass filtering, band-pass filtering (Franceschini et al., 2003; Jasdzewski et al., 2003; Cui, Bray, & Reiss, 2010), wavelet-based filtering (Jang et al., 2009; Lina et al., 2010; Matteau-Pelletier et al., 2009; Molavi & Dumont, 2012), correlation-based signal improvement (CBSI) (Cui, Bray, & Reiss, 2010), optimal filtering which are adaptive filtering (Abdelnour & Huppert, 2009; Zhang, Brown, & Strangman, 2007; M. Izzetoglu, K. Izzetoglu, Bunce, & Onaral, 2003), Wiener filtering (Izzetoglu, M., Izzetoglu, Bunce, & Onaral, 2003), Kalman filtering (Izzetoglu, Chitrapu, Bunce, & Onaral, 2010) and coefficient of variance related filtering which is Sliding-window Motion Artifact Rejection algorithm (SMAR) (Ayaz, Izzetoglu, Shewokis, & Onaral, 2010), that can effectively remove high frequency noises and low frequency drifts. Second category of artifact removal approaches is removing artifacts based on its spatial characteristics, which are eigenvector based algorithms that assumes signal due to noise is broadly spatially distributed compared to signal due to neural activity. Principal component analysis (PCA) and independent component analysis (ICA) based filtering algorithms are used for this purpose in many NIRS based studies (Ayaz, Izzetoglu, Shewokis, & Onaral, 2010; Zhang, Brooks, Franceschini, & Boas, 2005; Hiroyasu, Nakamura, & Yokouchi, 2013). However, PCA an ICA based filtering algorithms lead to make non-objective decisions about the number of components to keep, and they are challenging to apply to real-time processing of data. In the third category, artifacts are measured independently by additional hardware and excluded from the signal. For instance, head motion of the participants can be measured by using an accelerometer (Hiroyasu, Nakamura, & Yokouchi, 2013). Also, noise can be measured by using a channel with a very short distance between the emitter and detector in order to infrared light cannot pass through the brain tissue. The resultant signal must be noise, which can be subtracted from the signal (Cui, Bray, & Reiss, 2010).

In this thesis, in order to refine *f*NIR signals, firstly, low-pass with a finite impulse response (FIR), linear phase filter with an order 20 and cut-off frequency of 0.1 Hz was applied to raw *f*NIR signals in order to attenuate the high frequency signals and pass low frequency signals (Izzetoglu et al., 2005; Ayaz et al., 2011).

However, low-pass filtering cannot fully remove outliers or abrupt spikes in the data, which have amplitudes much higher or lower than regular optical signal values (Ayaz et al., 2010). To refine outliers and abrupt spikes from the *f*NIR signals, SMAR algorithm was applied to previously FIR filtered (refined) data by using default options. SMAR detects sudden local changes, which are much faster or slower than typical signals, and remove them from the data (Ayaz et al., 2010). After *f*NIR light data were filtered, task periods and local baselines were constructed. By using automated time markers, *f*NIR data epochs for 12 task periods (three trials for each n-back condition per participant) and 12 local baselines of related task periods were identified. After time synchronization markers were determined, refined *f*NIR light data were converted to hemodynamic responses by using previously mentioned modified Beer-Lambert Law to obtain oxy-hemoglobin, deoxy-hemoglobin, total hemoglobin, and difference in hemoglobin concentrations, oxygenation concentrations (Chance et al., 1993; Villringer and Chance, 1997; Ayaz, 2010). After refined light data was converted, hemodynamic data was extracted in blocks based on the conditions of the experiment, then data becomes ready to analyze, and visualize on topography.

CHAPTER 3

Results

3.1 Data Analysis Method

In social and behavioral sciences, most of the data have multilevel structure. In this type of data, which is also called 'hierarchically structured', measures are nested in different units. Multilevel data structures can be seen in repeated measures where an individual's responses over experimental conditions are correlated with each other. In repeated measures, which are obtained from a single individual at different conditions, the observations are nested within individuals. In other words, the bottom of the hierarchy, known as level-1, consists of within-individual components that are repeated measures obtained from the individuals. One level up of the hierarchy, known as level-2, includes between-individual components which are individuals. Level-1 component describes how each individual changes over repeated measurements; level-2 component describes how these changes vary across individuals (Field, Miles, & Field, 2012; Bryk & Raudenbush, 1987; 2012).

Multilevel models recognize the existence of such data hierarchies and allow investigating the relationships between variables in each level of the data simultaneously. Different names are used in the research literature referring the multilevel analysis of hierarchically structured (or multilevel structured) data (Kreft & De Leeuw, 1998; Raudenbush & Bryk, 2002; Hox, 2010) such as multilevel regression models, hierarchical linear models (HLM, Raudenbush & Bryk, 2002), random coefficient models (Longford, 1993), and mixed-effect models (Littell, et al., 2006). In this thesis, linear mixed effects model was preferred to be used. Although different names are used for this type of analysis, all of them are used for analyzing

hierarchical data, with one single outcome that is measured at the lowest level, and they allow including explanatory variables at all existing levels (Hox, 2010).

In the current study, hierarchical structure of the data has two levels; level-1 and level-2. Repeated measures obtained from different n-back conditions and trials were at level-1 and participants were level-2 in the hierarchy. Because both behavioral measurements and hemodynamic measures were nested within participants, multilevel linear mixed effects models were conducted in order to analyze the data.

Apart from the hierarchical structure of the data, there are several reasons to conduct multilevel linear mixed effects models analysis rather than conducting a conventional analysis, such as ANOVA, which is a commonly used method when repeated measures were obtained. Although conventional models were often applied in experimental studies, multilevel models are accepted as an alternative version of those conventional models due to its several advantages and solutions to problems that can be seen in ANOVA related to sphericity, hierarchical sampling, and missing data (Quene & Bergh, 2004; Hoffman, 2007; Hoffman & Rovine, 2007, Bryk & Raudenbush, 2002). In particular, sphericity assumption is tended to be violated when within-individual change is measured (Chan, 1998a; Kwok et al., 2007). If the sphericity assumption is not met, then tests of the fixed effects from the ANOVA model may be incorrect. Studies revealed MLMs enable to estimate a variety of alternative variance-covariance matrices and their fit can be compared in order to ensure the most appropriate tests of the fixed effects. Hence, MLMs can be used even if the assumption of sphericity is violated (Max & Onghena, 1999; Littell, Pendergast, & Natarajan, 2000; Quené & van den Bergh, 2004). Furthermore, conventional methods do not provide information to achieve more efficient statistical inferences, because they cannot be extended to higher-order sampling procedures. On the other hand, multilevel models allow to determine several different hierarchical models with varying complexity and to select the best fitting one (Gueorguieva & Krystal, 2004; Quené & van den Bergh, 2004). One of the advantages of MLMs over ANOVAs is that MLMs use all data on each individual and not affecting from randomly missing data. Data can be analyzed without any difficulties or special arrangement for missing cells. However, in ANOVA-based methods, only complete cases can be included in analysis which influences the power

of the experimental design (Kwok et al., 2009; Quené & van den Bergh, 2004). As this thesis includes brain imaging data, the inevitability of missing data in this type of data should be taken into consideration. Missing data occurs as a result of filtering abrupt changes in the signals, such as head and/or body movements of the participants, which are largely beyond the control of the experimenter. Because MLMs overcome with missing data compared to ANOVA, MLM is the appropriate analysis method to be used.

To describe the relationship between the outcome variable(s) and predictor variables, the distinction between the fixed effects and random effects is crucial (Onyango, 2009). Fixed effects are constant, or fixed, for all of the participants. As opposed to that, random effects vary among participants (Hoffman & Rovine, 2007).

In this thesis, n-back type (0-back, 1-back, 2-back and 3-back) and trials (trial 1, trial 2, and trial 3) are the level-1 and gender (male or female) is the level-2 explanatory variables introduced to the model to explain the variation on regression coefficients which are intercept and slope. These variables and their interactions in the model are the fixed effects of the mixed effect model. Random part of this model can be random intercept and random slope that can vary across participants. In random intercept model, as the name suggests intercepts are assumed to vary across individuals; but the effect of n-back type is, which is slope, assumed to be same. On the other hand, in random slope model participants are not only allowed to have varying intercepts, but they are also allowed to have different slopes showing the effect of n-back type. Those models were tested based upon starting from the simple (empty, or baseline) model and moving on the more complicated (full) models (Field, Miles, & Field, 2012). In the first model (empty model) no fixed effect with random intercept model was tested, and then fixed effects (n-back conditions, trials and gender) were added to random intercept and finally random slopes model with the fixed effects was tested.

In deciding which model to be used between two alternative models, the log likelihood value has a primary role (Healy, 2001). The log likelihood value indicates that how much unexplained information left after the model has been fitted. Thus, large values of the log likelihood statistic shows poor fit of statistical models, because if log likelihood values become larger, then this signifies there are more unexplained observations left. Therefore, in order to choose the model that best fits

the data; the log likelihoods of the two models were compared by looking at the difference between their deviances, which is usually denoted as (-2LL) and known as the likelihood ratio. The positive difference of 2LL has a chi-square (χ^2) distribution with degrees of freedom (*df*) obtained from the difference of the number of parameters to be estimated in the two models (Park & Lake, 2006; Field, Miles, & Field, 2012). In addition to that, the Akaike information criterion (AIC) (Akaike, 1977) and the Bayesian information criterion (BIC) (Schwarz, 1978) of the tested models which were also considered as likelihood based indices are considered (Hox, 2010; Field, Miles, & Field, 2012; Hoffman, 2007). If decreases in AIC, BIC and statistically significant log-likelihood ratios were obtained, then this shows the improvement of the model fit. If the difference between the log likelihood ratios of models is not statistically significant, then this implies that a more complex model may not be necessary to use (Hox, 2002). For example, if the random intercept and the random slope model were compared and if the χ^2 statistics have not been found statistically significant, then this means that the random intercept model was accepted as the best fitted model.

There are several statistical software packages that consist of analyzing methods for LMMs, such as SAS, SPSS, STATA, S+, and R. The main advantage of R is its free availability, dynamic development, open-source environment for statistical computing and graphics (Galecki & Burzykowski, 2013). Therefore, R version 3.3.2 statistical computing platform (www.r-project.org) was used for both analyzing behavioral and hemodynamic data. Multilevel linear mixed effects model were estimated using the lmer function of Linear Mixed-Effects Models by using lme4 package (Bates et al., 2013) and the statistical tests of the fixed effects parameters were calculated by using the lmerTest package in which the Satterthwaite approximation was implemented (Kuznetsova et. al., 2015).

For both behavioral and hemodynamic measurements, firstly the effect of adding fixed effects which are n-back type, trials and their interactions to the random intercept model was compared to the empty model to see whether the intercepts will vary across participants. If a statistically significant difference between log likelihood ratios was found, then the random slope model was tested in both behavioral measurements, which are the number of detected target letters, number of detected non-target letters, reaction time during target letter detection and reaction

time during non-target detection and hemodynamic measurements, which are oxy-Hb, deoxy-Hb, total-Hb, and oxygenation, in each of the 16-channels. After that, the construct validity of measurements was investigated. If statistically significant differences between log likelihood ratios of empty model, random intercept, and random slope model were found, then the fixed effect of gender was added to the statistically significant model in order to see whether both behavioral and hemodynamic measurements measure the same constructs in male and female participants.

Additionally, for the behavioral and hemodynamic measurements analyses, the intercept value was calculated from the mean values of the baseline levels of fixed effects. Therefore, 0-back condition in trial 1was accepted as the baseline variable or reference category for the behavioral and hemodynamic measurements analyses. Comparisons among the n-back conditions, trials, and their interactions were made according to the 0-back condition in trial 1. For the validity of behavioral and hemodynamic measurements, because gender was added as a fixed effect to the model, the intercept value was calculated as the mean value of 0-back condition, trial 1 and males. Hence, comparisons among the n-back conditions, trials, gender and their interactions were conducted based on the 0-back condition in trial 1 of males.

3.2 Behavioral Data Analysis

3.2.1 Number of Detected Target Letters

For the number of detected target letters, whether adding n-back conditions, trials and their interactions as fixed effects significantly improve the random intercept model was investigated. It was found that AIC declined from 2351.60 to 2111.20 and BIC reduced from 2364.60 to 2171.90, and the log likelihood ratio test indicated statistically significant difference between random intercept with n-back conditions, trials and their interactions model and empty model (χ^2 (11) = 262.40, p < .05). This means that including fixed effects of n-back conditions, trials and their interactions to the model result in significant improvement of the model. Hence model proceeded with random slope model.

It was observed that AIC decreased from 2351.60 to 2101.80 and BIC reduced from 2364.60 to 2201.50 and the log likelihood ratio test indicated statistically significant difference between random intercept and random slope model (χ^2 (9) =

27.36, $p < .05$). This result implied that random slope model also significantly improved the model.

Results of the random slope model indicated that the number of detected target letters was statistically significant under 0-back condition in trial 1, $b = 9.66$, $t(335.60) = 47.96$, $p < .05$. For the n-back conditions, 1-back condition, $b = -1.13$, $t(226.60) = -3.81$, $p < .05$; 2-back condition, $b = -1.70$, $t(167.60) = -5.42$, $p < .05$; and 3-back condition, $b = -3.32$, $t(171.60) = -10.27$, $p < .05$, significantly differed from 0-back condition in trial 1 in the number of detected target letters (Table 2). The negative coefficients of the main effect of n-back conditions reflected that the number of detected target letters significantly decreased in 1-, 2-, and 3-back conditions compared with the 0-back condition in trial 1 (Figure 7).

3.2.2 Number of Detected Non-Target Letters

For the number of detected non-target letters, whether adding n-back conditions, trials and their interactions as fixed effects significantly improve the random intercept model was investigated. It was observed that AIC declined from 2582.30 to 2215.90 and BIC reduced from 2595.30 to 2276.60, and the log likelihood ratio test indicated statistically significant difference between this model and empty model (χ^2 (11) = 388.39, $p < .05$), showing that including fixed effects of n-back conditions, trials and their interaction caused significant improvement of the model. Thus, model proceeded with random slope model.

It was observed that AIC decreased from 2582.30 to 2066.10 and BIC reduced from 2595.30 to 2165.80, and the log likelihood ratio test indicated statistically significant difference between random intercept and random slope model (χ^2 (9) = 167.80, $p < .05$). This result revealed that random slope model resulted in significant improvement of the model.

Results of the random slope model indicated that the number of detected non-target letters was statistically significant under 0-back condition in trial 1, $b = 9.66$, $t(335.60) = 47.96$, $p < .05$.

Table 2

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on the Number of Detected Target Letters

	b	*SE*	*df*	*t*	*p*
(Intercept)	9.66	0.20	335.60	47.96	2e-16 ***
1-back	-1.13	0.30	226.60	-3.81	0.00 ***
2-back	-1.70	0.31	167.60	-5.42	2.03e-07 ***
3-back	-3.32	0.32	171.60	-10.27	2e-16 ***
Trial2	-0.04	0.28	437.40	-0.15	0.88
Trial3	-0.15	0.28	437.40	-0.53	0.59
1-back:Trial2	0.43	0.39	437.40	1.08	0.28
2-back:Trial2	-0.17	0.39	437.40	-0.43	0.67
3-back:Trial2	0.57	0.39	437.40	1.46	0.15
1-back:Trial3	0.32	0.39	437.40	0.81	0.42
2-back:Trial3	-0.11	0.39	437.40	-0.27	0.79
3-back:Trial3	0.26	0.39	437.40	0.65	0.52

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

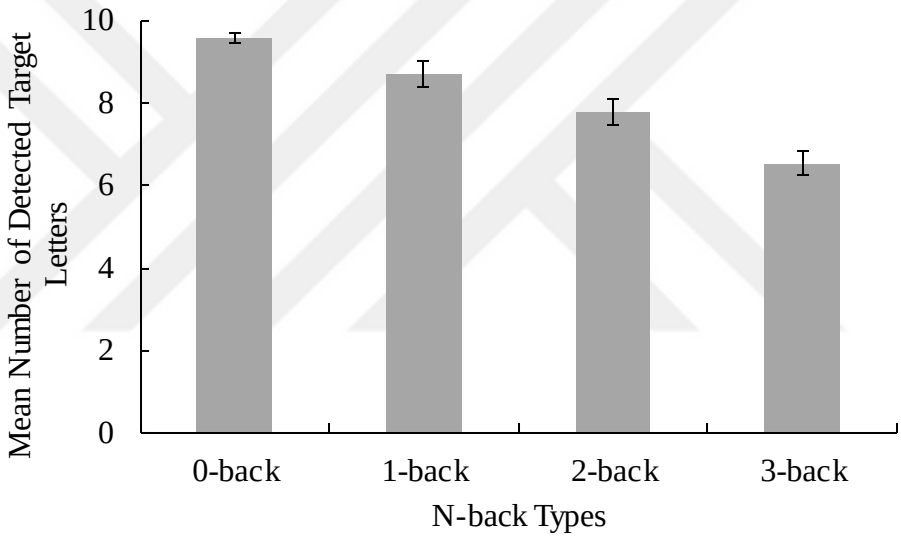

Figure 7. Mean number of detected target letters for each n-back condition (Error bars indicate 95% Confidence Interval).

Compared with 0-back in trial 1, the number of detected non-target letters significantly varied from 1-back condition, b = -1.13, t(257.00) = -4.15, p < .05; 2-back condition, b = -2.38, t(111.00) = -6.86, p < .05; and 3-back condition, b = -4.89, t(92.80) = -12.82, p < .05. The negative coefficients indicated significant decrease in the number of detected non-target letters in 1-, 2-, and 3-back conditions to the compared with the 0-back condition in trial 1 (Table 3).

In addition, the interaction between n-back conditions and trials was also found statistically significant. Compared with the 0-back condition in trial 1, number of detected non-target letters in 3-back in trial 2 and trial 3 significantly varied, (b = 0.94, t(517.00) = 2.05, p < .05; b = 1.45, t(517.00) = 3.16, p < .05, respectively). For the interaction effect of n-back condition and trial, the positive coefficients indicated significant increase in the number of detected non-target letters in 3-back in trial 2 and trial 3, compared with 0-back condition in trial 1 (Figure 8).

3.2.3 Reaction Time During Target Letter Detection

For the reaction time during target letter detection, similar to previous model testing, whether adding n-back conditions, trials and their interactions as fixed effects significantly improve the random intercept model was examined. It was found that AIC declined from -45.32 to -316.45 and BIC reduced from -32.31 to -255.76, and the log likelihood ratio test indicated statistically significant difference between random intercept with n-back conditions, trials and their interactions model and empty model (χ^2 (11) = 293.13, p < .05). This means that including fixed effects of n-back conditions, trials and their interaction leads significant improvement of the model. Therefore, model proceeded with random slope model.

It was observed that AIC decreased from -45.32 to -410.41 and BIC reduced from -32.31 to -310.70, and the log likelihood ratio test indicated statistically significant difference between random intercept and random slope models (χ^2 (9) = 111.96, p < .05). This result indicated that the random slope model also significantly improved the model.

Table 3

Multilevel Analyses of the Effect of N-back Conditions, Trial and Their Interaction on the Number of Non-Target Letters

	b	*SE*	*df*	*t*	*p*
(Intercept)	19.83	0.18	415.10	111.25	2e-16 ***
1-back	-1.13	0.27	257.00	-4.15	4.55e-05 ***
2-back	-2.38	0.35	111.00	-6.86	4.07e-10 ***
3-back	-4.89	0.38	92.80	-12.82	2e-16 ***
Trial2	-0.13	0.25	423.00	-0.51	0.61
Trial3	-0.15	0.25	423.00	-0.59	0.55
1-back:Trial2	0.45	0.36	423.00	1.26	0.21
2-back:Trial2	0.26	0.36	423.00	0.72	0.47
3-back:Trial2	0.94	0.36	423.00	2.63	0.01 **
1-back:Trial3	0.68	0.36	423.00	1.92	0.42
2-back:Trial3	0.43	0.36	423.00	1.20	0.79
3-back:Trial3	1.45	0.36	423.00	4.07	5.63e-05 ***

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

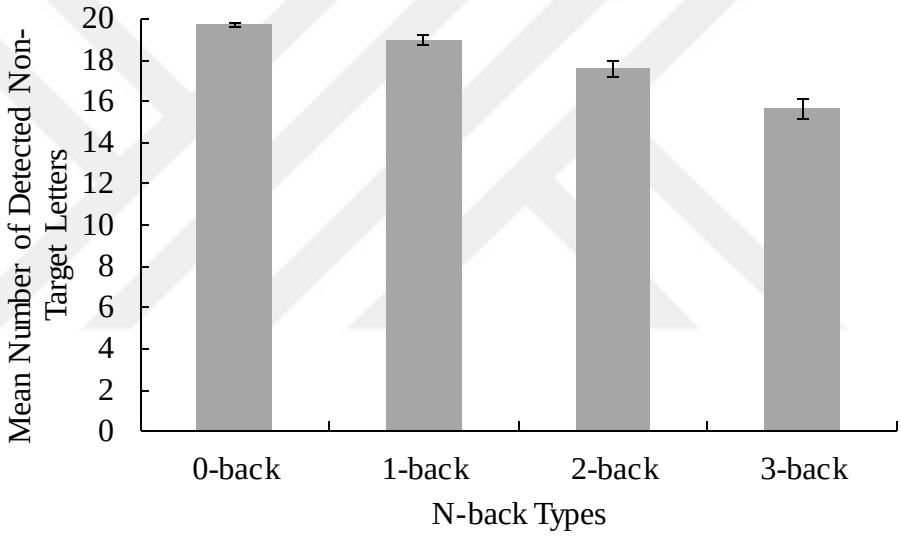

Figure 8. Mean number of detected non-target letters for each n-back condition (Error bars indicate 95% Confidence Interval).

Results of the random slope model indicated significant intercept, $b = 0.52$, $t(146.90) = 21.66$, $p < .05$. Compared with the 0-back condition in trial 1, 1-back condition, $b = 0.15$, $t(221.30) = 4.77$, $p < .05$; 2-back condition, $b = 0.24$, $t(151.60) = 7.30$, $p < .05$; and 3-back condition, $b = 0.41$, $t(107.60) = 10.76$, $p < .05$ significantly varied in reaction time during target letter detection. The positive coefficients reflected significant increase in the reaction time during target letter detection in 1-, 2-, and 3-back conditions compared with the 0-back condition in trial 1 (Table 4).

Moreover, the interaction effect between n-back conditions and trials on reaction time during target letter detection was found statistically significant. Reaction time during target letter detection significantly differed in 3-back in trial 3, compared with the 0-back condition in trial 1 ($b = -0.10$, $t(423.00) = -2.49$, $p < .05$). For the interaction effect of n-back condition and trial, the negative coefficient showed significant decrease in reaction time during target letter detection of 3-back in trial 3 compared with the 0-back condition trial 1 (Table 4 & Figure 9).

3.2.4 Reaction Time During Non-Target Letter Detection

For the reaction time during non-target letter detection, whether adding n-back conditions, trials and their interactions as fixed effects significantly improve the random intercept model was examined. It was observed that AIC declined from -49.25 to -435.55 and BIC reduced from -36.25 to -374.86, and the log likelihood ratio test indicated statistically significant difference between random intercept with n-back conditions, trials and their interactions model and empty model (χ^2 (11) = 408.30, $p < .05$). This result means that including fixed effects of n-back conditions, trials and their interaction result in significant improvement of the model and thus model proceeded with random slope model.

It was observed that AIC decreased from -49.25 to -552.64 and BIC was reduced from -36.25 to -452.93 and the log likelihood ratio test indicated statistically significant difference between random intercept and random slope model (χ^2 (9) = 135.09, $p < .05$). This result revealed that random slope model also significantly improved the model.

Table 4

Multilevel Analyses of the Effect of N-back Conditions on the Reaction Time During Target Letter Detection

	b	*SE*	*df*	*t*	*p*
(Intercept)	0.52	0.02	146.90	21.66	2e-16 ***
1-back	0.15	0.03	211.30	4.77	3.45e-06 ***
2-back	0.24	0.03	151.60	7.30	1.48e-11 ***
3-back	0.41	0.04	107.60	10.76	2e-16 ***
Trial2	0.01	0.03	423.00	0.52	0.60
Trial3	-0.01	0.03	423.00	-0.19	0.85
1-back:Trial2	-0.06	0.04	423.00	-1.55	0.12
2-back:Trial2	-0.06	0.04	423.00	-1.44	0.15
3-back:Trial2	-0.07	0.04	423.00	-1.73	0.09
1-back:Trial3	-0.05	0.04	423.00	-1.31	0.19
2-back:Trial3	-0.04	0.04	423.00	-0.95	0.34
3-back:Trial3	-0.10	0.04	423.00	-2.49	0.01 *

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

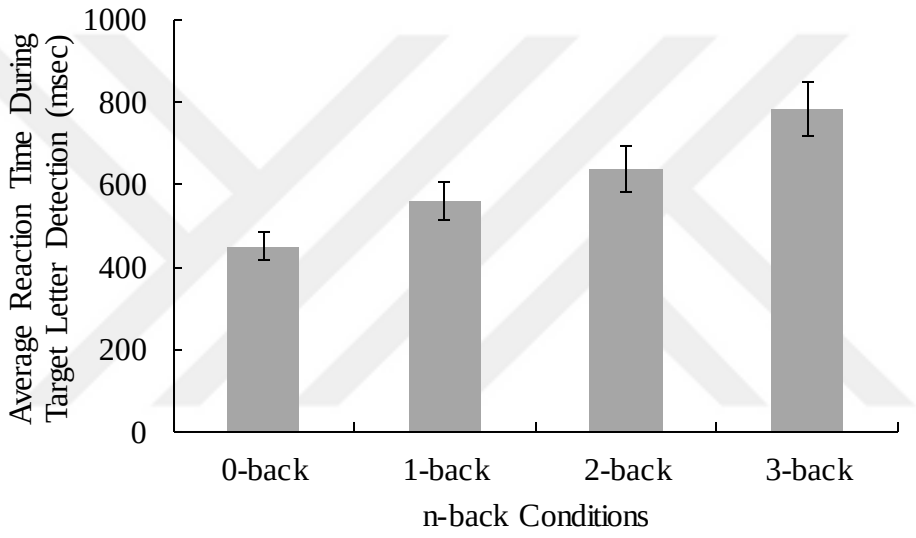

Figure 9. Mean reaction time during target letter detection for each n-back Condition (Error bars indicate 95% Confidence Interval).

Results of the random slope model showed significant intercept, $b = 0.53$, $t(105.10) = 23.00$, $p < .05$. For the n-back conditions, 1-back condition, $b = 0.19$, $t(146.40) = 6.69$, $p < .05$; 2-back condition, $b = 0.33$, $t(116.70) = 10.67$, $p < .05$; and 3-back condition, $b = 0.42$, $t(102.00) = 12.54$, $p < .05$, significantly differed from 0-back condition in trial 1 in the reaction time during non-target letter detection. The positive coefficients demonstrated significant increase in the reaction time during non-target letter detection in 1-, 2-, and 3-back conditions compared with the 0-back condition in trial 1.

Furthermore, the interaction effect between n-back conditions and trials on reaction time during non-target letter detection was found statistically significant. As in the reaction time during target letter detection, the reaction time during non-target letter detection was significantly varied in 3-back in trial 3, compared with the 0-back condition in trial 1 ($b = -0.07$, $t(423.00) = -2.11$, $p < .05$) (Table 5). The negative coefficients showed significant decrease in reaction time during non- target letter detection of 3-back in trial 3 compared with the 0-back condition in trial 1 (Figure 10).

3.3 Hemodynamic Activity Data Analysis

In this part, hemodynamic findings will be presented for each measurement separately.

3.3.1 Oxy-Hb Measurements

For the oxy-Hb measures, whether adding n-back conditions, trials and their interactions as fixed effects significantly improve the model was investigated.

It was found that in the following channels AICs decreased after random intercept model was tested and the log likelihood ratio tests indicated statistically significant difference between random intercept and empty model: in Channel 2 AICs decreased from 1232.50 to 1215.80 and the log likelihood ratio test was statistically significant (χ^2 (11) = 38.80, $p < .05$); in Channel 4 the decrease was from 1246.30 to 1246.20 and the log likelihood ratio test was significant (χ^2 (11) = 22.08, $p < .05$); in Channel 6 it was from 1227.90 to 1226.70 and the log likelihood ratio test was again significant (χ^2 (11) = 23.22, $p < .05$), except for Channel 16 in which AICs remained identical but the log likelihood ratio test was significant (χ^2 (11) = 20.94, $p < .05$). So, model proceeded with random slope model for those channels.

Table 5

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on the Reaction Time During Non-Target Letter Detection

	b	*SE*	*df*	*t*	*p*
(Intercept)	0.53	0.02	105.10	23.00	2e-16 ***
1-back	0.19	0.03	146.40	6.69	4.32e-10 ***
2-back	0.33	0.03	116.70	10.67	2e-16 ***
3-back	0.42	0.03	102.00	12.54	2e-16 ***
Trial2	0.01	0.02	423.00	0.32	0.75
Trial3	-0.02	0.02	423.00	-0.72	0.47
1-back:Trial2	-0.04	0.03	423.00	-1.24	0.22
2-back:Trial2	-0.03	0.03	423.00	-0.77	0.44
3-back:Trial2	-0.03	0.03	423.00	-1.03	0.31
1-back:Trial3	-0.04	0.03	423.00	-1.14	0.26
2-back:Trial3	-0.04	0.03	423.00	-1.20	0.23
3-back:Trial3	-0.07	0.03	423.00	-2.11	0.04 *

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

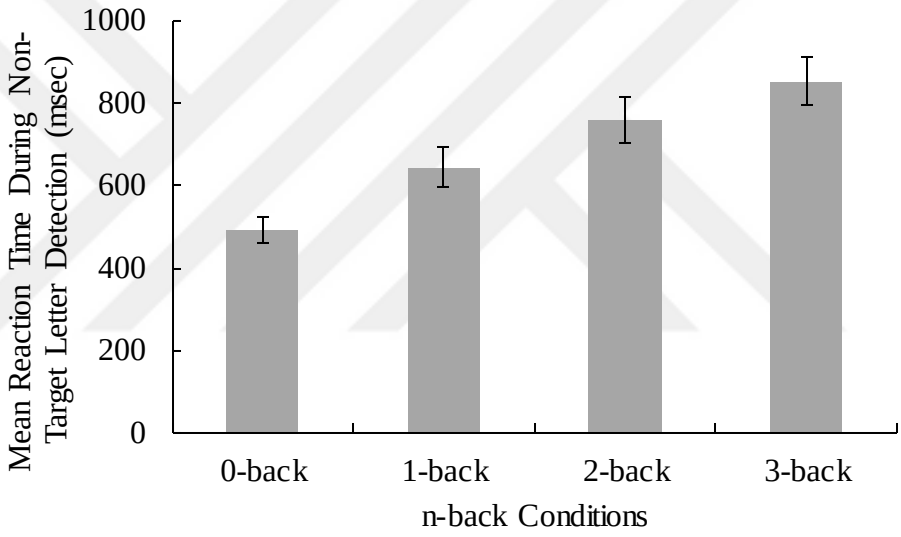

Figure 10. Mean reaction time during non-target letter detection for each n-back condition (Error bars indicate 95% Confidence Interval).

However, adding n-back conditions, trials and their interactions as fixed effects did not significantly improve the model in Channel 1 (χ^2 (11) = 17.29, p > .05), Channel 3 (χ^2 (11) = 13.52, p > .05), Channel 5 (χ^2 (11) = 11.62, p > .05), Channel 7 (χ^2 (11) = 16.34, p > .05), Channel 8 (χ^2 (11) = 15.94, p > .05), Channel 9 (χ^2 (11) = 11.92, p > .05), Channel 10 (χ^2 (11) = 9.28, p > .05), Channel 11 (χ^2 (11) = 11.35, p > .05), Channel 12 (χ^2 (11) = 14.73, p > .05), Channel 13 (χ^2 (11) = 10.04, p > .05), Channel 14 (χ^2 (11) = 16.87, p > .05), Channel 15 (χ^2 (11) = 17.43, p > .05). Therefore, other models were not tested across those channels.

Due to their statistically significance in the first model comparison, random slope model was tested in Channel 2, Channel 4, Channel 6 and Channel 16. Results indicated that the random slope model did not improve the model in Channel 2 (χ^2 (9) = 7.10, p > .05), Channel 4 (χ^2 (9) = 16.39, p > .05), Channel 6 (χ^2 (9) = 11.40, p > .05). Therefore, random intercept model was accepted as the best fitted model for those channels. As distinct from those channels, random slope model significantly improve the model in Channel 16. It was observed that in Channel 16, AIC decreased from 1389.00 to 1385.40 and the log likelihood ratio test showed statistically significant difference between random intercept and random slope models (χ^2 (9) = 22.68, p < .05). This result implied that random slope model should be accepted as the best fitted model for Channel 16.

Because significant activation in each of the 16-channels from the same models was separately tested for each participant, Benjamini–Hochberg corrections with q specified at 0.05 were applied. Correction was applied to the significance values of each 16-channel, in order to control the false discovery error rate (FDR) (Benjamini & Hochberg, 1995). After FDR correction was applied to all 16-channels, statistically significance was found only in Channel 2.

For Channel 2, results of the random intercept model indicated that oxy-Hb measurements significantly differed in the 0-back condition in trial 1 from 3-back condition (b = -0.35, t(493.60) = -2.35, p < .05). The negative coefficient signified significant decrease in oxy-Hb measurements in 3-back condition compared with the 0-back condition in trial 1. In addition, oxy-Hb measurements significantly differed in the 0-back condition in trial 1 from trial 2 and trial 3 (b = –0.51, t(494.50) = –3.34, p < .01; b = –0.37, t(494.60) = –2.48, p < .01, respectively). The negative

coefficients showed significant decrease in oxy-Hb measurements in 3-back condition, compared with the 0-back condition in trial 1.

Furthermore, the interaction between 3-back in trial 2 compared with the 0-back condition in trial 1 was found statistically significant. Indicating that oxy-Hb measurements significantly differed in 3-back condition in trial 2 ($b = 0.56$, $t(494.00) = 2.62$, $p < .05$) from 0-back condition in trial 1. The positive coefficient of the interaction signified statistically significant increase in oxy-Hb measures in 3-back in trial 2, compared with the 0-back condition in trial 1 (Table 6). Visualization of the oxy-Hb activity of the prefrontal cortex was carried out by averaging oxy-Hb measurements based upon three trials and four n-back conditions (Figure 11).

3.3.2 Deoxy-Hb Measurements

For the deoxy-Hb measures, whether adding n-back conditions, trials and their interactions as fixed effects significantly improve the random intercept model was examined.

It was observed that in Channel 1, AIC decreased from 408.06 to 403.55 and the log likelihood ratio test was statistically significant (χ^2 (11) = 26.50, $p < .05$); in Channel 2, it decreased from 585.74 to 584.52 and log likelihood ratio test was significant (χ^2 (11) = 23.22, $p < .05$); in Channel 4, AIC declined from 379.04 to 370.85 and the log likelihood ratio test was significant (χ^2 (11) = 30.20, $p < .05$); in Channel 6, AIC decreased from 316.72 to 314.65 and the log likelihood ratio test was again significant (χ^2 (11) = 24.07 $p < .05$); in Channel 12, it was from 268.57 to 256.10 and the log likelihood ratio test was again significant (χ^2 (11) = 34.47 $p < .05$); in Channel 15, decreased from 218.81 to 206.39 and the log likelihood ratio test was significant (χ^2 (11) = 34.43 $p < .05$); in Channel 16, AIC declined from 496.44 to 487.29 and the log likelihood ratio test again was statistically significant (χ^2 (11) = 31.15 $p < .05$), except for Channel 8, Channel 9, Channel 11, Channel 14, AICs remained same, but the log likelihood ratio tests were significant (χ^2 (11) = 21.92 $p < .05$; χ^2 (11) = 21.61 $p < .05$; χ^2 (11) = 21.97 $p < .05$, χ^2 (11) = 21.73 $p < .05$, respectively). Those findings demonstrated that including fixed effects of n-back conditions, trials and their interactions significantly improved the model.

Table 6

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on oxy-Hb Measures in Channel 2

	b	*SE*	*df*	*t*	*p*
(Intercept)	0.17	0.11	535.00	1.56	0.12
1-back	0.08	0.15	494.50	0.55	0.58
2-back	-0.08	0.15	494.50	-0.54	0.59
3-back	-0.35	0.15	493.60	-2.35	0.02 *
Trial2	-0.51	0.15	494.50	-3.34	0.00 ***
Trial3	-0.37	0.15	494.60	-2.48	0.01 *
1-back:Trial2	-0.15	0.22	494.00	-0.69	0.49
2-back:Trial2	0.35	0.21	495.80	1.65	0.10
3-back:Trial2	0.56	0.21	494.00	2.62	0.01 **
1-back:Trial3	-0.14	0.21	494.00	-0.68	0.50
2-back:Trial3	0.04	0.21	494.00	0.17	0.87
3-back:Trial3	0.28	0.21	493.60	1.33	0.18

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

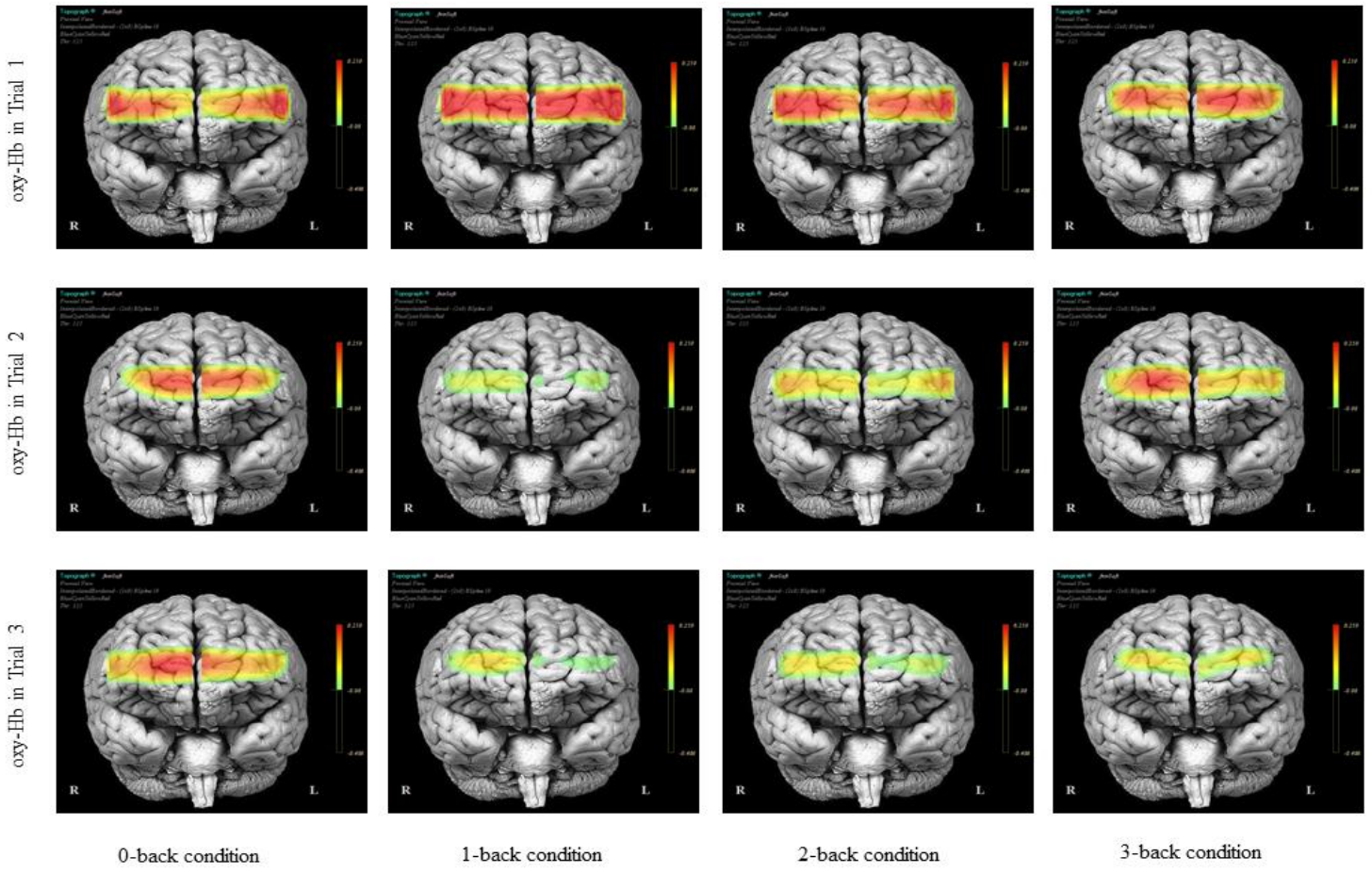

Figure 11. Mean oxy-Hb measurements based on three trials and four n-back conditions.

However, model was not improved in Channel 3 (χ^2 (11) = 15.81, p > .05), Channel 5 (χ^2 (11) = 19.70, p > .05), Channel 7 (χ^2 (11) = 18.80, p > .05), Channel 10 (χ^2 (11) = 19.31, p > .05), Channel 13 (χ^2 (11) = 18.77, p > .05). Thus, other models were not tested among those channels.

Random slope model was tested in Channel 1, Channel 2, Channel 4, Channel 6, Channel 8, Channel 9, Channel 11, Channel 12, Channel 14, Channel 15, and Channel 16 because of their statistical significance in the first model comparison. Results showed that random slopes model was significant in the following channels: in Channel 1, AIC decreased from 408.06 to 378.00 and the log likelihood ratio test was statistically significant (χ^2 (9) = 43.56, p < .05); in Channel 2, it was from 585.74 to 575.19 and the log likelihood ratio test was again significant (χ^2 (9) = 27.33, p < .05); in Channel 9, declined from 463.13 to 450.07 and the log likelihood ratio test was statistically significant (χ^2 (9) = 31.45, p < .05). These results indicated that random slope model also significantly improved the model.

On the other hand, random slope model did not improve the model in the following channels: in Channel 4 (χ^2 (9) = 9.13, p > .05); Channel 6 (χ^2 (9) = 11.60, p > .05); Channel 8 (χ^2 (9) = 17.19, p > .05); Channel 11 (χ^2 (9) = 4.96, p > .05); Channel 12 (χ^2 (9) = 10.81, p > .05); Channel 14 (χ^2 (9) = 10.05, p > .05); Channel 15 (χ^2 (9) = 15.84, p > .05); and Channel 16 (χ^2 (9) = 6.58, p > .05). Therefore, random intercept model was accepted as the best fitted model for those channels.

After FDR correction was applied to all 16-channels, statistically significant effect of n-back condition, trial and their interaction on deoxy-Hb measures was found in Channel 1, 2, 4, 6, 8, 9, 11, 12, 14, 15, and 16.

As it was stated above, only the random intercept with n-back conditions, trials and their interactions model significantly improved the empty model in Channel 4, Channel 6, Channel 8, Channel 11, Channel 12, Channel 14, Channel 15, and Channel 16.

For Channel 4, results of the random intercept model indicated that deoxy-Hb measurements significantly differed in trial 2 and trial 3 (b = 0.21, t(517.00) = 3.20, p < .01; b = 0.25, t(517.00) = 3.83, p < .001, respectively) from the 0-back condition in trial 1. The positive coefficients of trial 2 and trial 3 showed significant increases in deoxy-Hb compared with the 0-back condition in trial 1. Furthermore, the

interactions between 2-back in trial 2 and trial 3 and the interaction between 3-back in trial 3, compared with the 0-back condition in trial 1 were found statistically significant. The interactions revealed that deoxy-Hb measurements significantly differed in 2-back condition in trial 2 (b = -0.19, t(517.00) = -2.05, p < .05), and 2-back condition in trial 3 (b = -0.28, t(517.00) = -3.03, p < .01) and in 3-back in trial 3 (b = -0.32, t(517.00) = -3.43, p < .001), compared with the 0-back condition in trial 1. The negative coefficients of the interactions showed statistically significant decreases in deoxy-Hb measurements in 2-back in trial 2 and trial 3, also 3-back in trial 3 compared with the 0-back condition in trial 1 (Table 7).

In Channel 6, results of the random intercept model showed that deoxy-Hb measurements were significantly different from trial 2 and trial 3 (b = 0.14, t(517.00) = 2.22, p < .05; b = 0.25, t(517.00) = 3.83, p < .001, respectively), compared with the 0-back condition in trial 1. As in Channel 4, the positive coefficients revealed significant increase in deoxy-Hb measurements, compared with the 0-back condition in trial 1. In addition, interactions revealed that deoxy-Hb measurements significantly differed in 2-back in trial 2 and trial 3 (b = -0.20, t(517.00) = -2.28, p < .05; b = -0.23, t(517.00) = -2.63, p < .01) and in 3-back in trial 3 (b = -0.26, t(517.00) = -2.99, p < .01) from 0-back condition in trial 1. The negative coefficients of the interactions indicated statistically significant decrease in deoxy-Hb measurements in 2-back in trial 2 and trial 3 and 3-back in trial 2, compared with the 0-back condition in trial 1 (Table 8).

In Channel 8, results of the random intercept model indicated that deoxy-Hb measurements significantly varied in trial 2 compared with the 0-back condition in trial 1 (b = 0.27, t(514.10) = 3.26, p < .05). The positive coefficient of trial 2 showed significant increase in deoxy-Hb compared with the 0-back condition in trial 1. The interactions indicated that deoxy-Hb measurements significantly differed in 2-back in trial 2 and trial 3, from 0-back condition in trial 1 (b = -0.28, t(514.10) = -2.33, p < .05; b = -0.24, t(514.40) = -2.04, p < .05, respectively). The negative coefficients of the interactions showed statistically significant decrease in deoxy-Hb measurements in 2-back in trial 2 and trial 3, compared with the 0-back condition in trial 1 (Table 9).

Table 7

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on deoxy-Hb Measurement in Channel 4

	b	*SE*	*df*	*t*	*p*
(Intercept)	-0.13	0.05	534.00	-2.79	0.01 **
1-back	0.04	0.07	517.00	0.61	0.54
2-back	0.07	0.07	517.00	1.10	0.27
3-back	0.12	0.07	517.00	1.75	0.08
Trial2	0.21	0.07	517.00	3.20	0.00 **
Trial3	0.25	0.07	517.00	3.83	0.00 ***
1-back:Trial2	-0.16	0.09	517.00	-1.74	0.08
2-back:Trial2	-0.19	0.09	517.00	-2.05	0.04 *
3-back:Trial2	-0.16	0.09	517.00	-1.67	0.10
1-back:Trial3	-0.10	0.09	517.00	-1.13	0.26
2-back:Trial3	-0.28	0.09	517.00	-3.03	0.00 **
3-back:Trial3	-0.32	0.09	517.00	-3.43	0.00 ***

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Table 8

Multilevel Analyses of the Effect of N-back Condition on deoxy-Hb Measurement in Channel 6

	b	*SE*	*df*	*t*	*p*
(Intercept)	-0.01	0.05	477.40	-0.11	0.91
1-back	-0.02	0.06	517.00	-0.37	0.72
2-back	0.03	0.06	517.00	0.52	0.60
3-back	0.08	0.06	517.00	1.27	0.21
Trial2	0.14	0.06	517.00	2.22	0.03 *
Trial3	0.14	0.06	517.00	2.29	0.02 *
1-back:Trial2	-0.09	0.09	517.00	-1.06	0.29
2-back:Trial2	-0.20	0.09	517.00	-2.28	0.02 *
3-back:Trial2	-0.17	0.09	517.00	-1.90	0.06
1-back:Trial3	-0.07	0.09	517.00	-0.77	0.44
2-back:Trial3	-0.23	0.09	517.00	-2.63	0.01 **
3-back:Trial3	-0.26	0.09	517.00	-2.99	0.00 **

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Table 9

Multilevel Analyses of the Effect of N-back condition on deoxy-Hb Measurement in Channel 8

	b	*SE*	*df*	*t*	*p*
(Intercept)	-0.04	0.07	402.40	-0.66	0.51
1-back	0.02	0.08	514.10	0.26	0.79
2-back	0.10	0.08	514.10	1.22	0.22
3-back	0.16	0.08	514.10	1.95	0.05
Trial2	0.27	0.08	514.10	3.26	0.00 **
Trial3	-0.15	0.08	514.70	1.65	0.10
1-back:Trial2	-0.19	0.12	514.10	-1.62	0.11
2-back:Trial2	-0.28	0.12	514.10	-2.33	0.02 *
3-back:Trial2	-0.22	0.12	514.10	-1.84	0.07
1-back:Trial3	-0.05	0.12	514.40	-0.38	0.70
2-back:Trial3	-0.24	0.12	514.40	-2.04	0.04 *
3-back:Trial3	-0.12	0.12	514.30	-1.04	0.30

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

In Channel 11, the results of the random intercept model only showed statistically significant difference in trial 2 on deoxy-Hb measurements when compared with the 0-back condition in trial 1 (b = 0.15, t(467.30) = 2.12, p < .05). The positive coefficient of trial 2 showed significant increase in deoxy-Hb compared with the 0-back condition in trial 1 (Table 10).

In Channel 12, results of the random intercept model demonstrated similar results with Channel 6. Compared with the 0-back condition in trial 1, deoxy-Hb measurements were significantly differed in trial 2 and trial 3 (b = 0.18, t(516.00) = 3.08, p < .001; b = 0.17, t(516.30) = 2.87, p < .001, respectively). The positive coefficients of trial 2 and trial 3 showed significant increase in deoxy-Hb measurements when compared with the 0-back condition in trial 1. The interactions between 2-back in trial 2 and trial 3, (b = -0.24, t(516.00) = -2.90, p < .001; b = -0.21, t(516.20) = -2.59, p < .01, respectively), and the interaction between 3-back in trial 3 (b = -0.20, t(516.20) = -2.45, p < .05) were statistically varied from 0-back condition in trial 1 on deoxy-Hb measurements. The negative coefficients of the interactions indicated statistically significant decrease in deoxy-Hb in 2-back in trial 2 and trail 3, 3-back in trial 3 when compared with the 0-back condition in trial 1 (Table 11).

In Channel 14, results of the random intercept model indicated that compared to 0-back condition in trial 1, deoxy-Hb measurements significantly differed in trial 2 and trial 3 (b = 0.16, t(514.90) = 2.28, p < .05; b = 0.16, t(514.90) = 2.27, p < .05, respectively). The positive coefficients of trial 2 and trial 3 indicated significant increase in deoxy-Hb measurements compared with the 0-back condition in trial 1.The interaction between 3-back in trial 3 significantly differed from the 0-back condition in trial 1. The interaction revealed that deoxy-Hb measurements significantly differed in 0-back condition in trial 1 from 3-back condition in trial 3 (b = -0.20, t(514.60) = -1.99, p < .05). The negative coefficient of the interaction indicated statistically significant decrease in deoxy-Hb measurements in 3-back in trial 3, compared with the 0-back condition in trial 1 (Table 12).

Table 10

Multilevel Analyses of the Effect of N-back condition on deoxy-Hb Measurement in Channel 11

	b	*SE*	*df*	*t*	*p*
(Intercept)	-0.02	0.05	252.00	-0.31	0.76
1-back	-0.04	0.07	284.00	-0.52	0.60
2-back	0.02	0.07	402.20	0.29	0.77
3-back	0.00	0.07	237.00	-0.02	0.98
Trial2	0.15	0.07	467.30	2.12	0.03 *
Trial3	0.12	0.07	469.70	1.69	0.09
1-back:Trial2	0.02	0.10	467.30	0.20	0.85
2-back:Trial2	-0.18	0.10	467.30	-1.83	0.07
3-back:Trial2	0.03	0.10	467.30	0.29	0.77
1-back:Trial3	0.04	0.10	468.50	0.38	0.71
2-back:Trial3	-0.11	0.10	468.50	-1.13	0.26
3-back:Trial3	-0.07	0.10	469.50	-0.75	0.46

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Table 11

Multilevel Analyses of the Effect of N-back condition, Trial and Their Interaction on deoxy-Hb Measurement in Channel 12

	b	*SE*	*df*	*t*	*p*
(Intercept)	-0.02	0.05	394.00	-0.40	0.69
1-back	-0.04	0.06	516.00	-0.69	0.49
2-back	0.04	0.06	516.00	0.76	0.45
3-back	0.05	0.06	516.00	0.78	0.43
Trial2	0.18	0.06	516.00	3.08	0.00 **
Trial3	0.17	0.06	516.30	2.87	0.00 **
1-back:Trial2	-0.04	0.08	516.00	-0.51	0.61
2-back:Trial2	-0.24	0.08	516.00	-2.90	0.00 **
3-back:Trial2	-0.08	0.08	516.00	-1.03	0.31
1-back:Trial3	-0.03	0.08	516.00	-0.41	0.68
2-back:Trial3	-0.21	0.08	516.20	-2.59	0.01 **
3-back:Trial3	-0.20	0.08	516.20	-2.45	0.01 *

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Table 12

Multilevel Analyses of the Effect of N-back condition, Trial and Their Interaction on deoxy-Hb Measurement in Channel 14

	b	*SE*	*df*	*t*	*p*
(Intercept)	-0.05	0.05	552.00	-0.92	0.36
1-back	-0.09	0.07	514.20	-1.30	0.20
2-back	0.04	0.07	514.90	0.60	0.55
3-back	0.04	0.07	514.90	0.54	0.59
Trial2	0.16	0.07	514.90	2.28	0.02 *
Trial3	0.16	0.07	514.90	2.27	0.02 *
1-back:Trial2	-0.02	0.10	514.20	-0.21	0.83
2-back:Trial2	-0.14	0.10	514.60	-1.37	0.17
3-back:Trial2	-0.12	0.10	514.90	-1.23	0.22
1-back:Trial3	0.04	0.10	514.20	0.38	0.70
2-back:Trial3	-0.11	0.10	514.60	-1.07	0.28
3-back:Trial3	-0.20	0.10	514.60	-1.99	0.047 *

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

In Channel 15, results of the random intercept model revealed significant intercept (b = -0.15, t(328.90) = -3.05, p < .01). The negative coefficient of 0-back condition in trial 1 showed significant decrease in deoxy-Hb measurements. Compared with the 0-back condition in trial 1, deoxy-Hb measurements significantly differed in 2-back and 3-back conditions (b = 0.17, t(417.00) = 2.60, p < .01; b = 0.17, t(417.00) = 2.65, p < .01, respectively) and in trial 2 and trial 3 (b = 0.25, t(417.00) = 3.91, p < .001; b = 0.29, t(417.30) = 4.48, p < .001, respectively). The positive coefficients reflected significant increases in deoxy-Hb measurements.

Furthermore, the interactions revealed that compared with the 0-back condition in trial 1, 2-back condition in trial 2 and trial 3 (b = -0.28, t(417.00) = -3.13, p < .01; b = -0.34, t(417.20) = -3.71, p < .001, respectively) and 3-back condition in trial 2 and trial 3 (b = -0.28, t(417.00) = -3.10, p < .01; b = -0.31, t(417.20) = -3.40, p < .001, respectively) statistically differed on deoxy-Hb measures. The negative coefficient of the interaction effects indicated statistically significant decrease in deoxy-Hb measures (Table 13).

In Channel 16, results of the random intercept model indicated significant intercept (b = -0.12, t(442.00) = -2.17, p < .05). The negative coefficient of 0-back condition in trial 1 showed significant decline in deoxy-Hb measurements. Compared with the 0-back condition in trial 1, deoxy-Hb measurements significantly differed in 3-back condition (b = 0.15, t(505.10) = 2.03, p < .05). The positive coefficient indicated significant increases in deoxy-Hb measurements

Moreover, deoxy-Hb measurements significantly varied in trial 2 and trial 3 (b = 0.26, t(505.10) = 3.57, p < .001; b = 0.28, t(505.10) = 3.87, p < .001, respectively), when compared with the 0-back condition in trial 1. The positive coefficients reflected significant increases in deoxy-Hb measurements. Furthermore, the interactions indicated significant differences in 3-back condition in trial 2 and trial 3, compared with the 0-back condition in trial 1 (b = -0.29, t(505.20) = -2.76, p < .01; b = -0.37, t(505.10) = -3.57, p < .001, respectively). The negative coefficients of the interactions indicated statistically significant decrease in deoxy-Hb measurements (Table 14).

Table 13

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on deoxy-Hb Measurement in Channel 15

	b	*SE*	*df*	*t*	*p*
(Intercept)	-0.15	0.05	328.90	-3.05	0.00 **
1-back	0.01	0.06	417.00	0.12	0.91
2-back	0.17	0.06	417.00	2.60	0.01 **
3-back	0.17	0.06	417.00	2.65	0.01 **
Trial2	0.25	0.06	417.00	3.91	0.00 ***
Trial3	0.29	0.06	417.30	4.48	0.00 ***
1-back:Trial2	-0.11	0.09	417.00	-1.22	0.22
2-back:Trial2	-0.28	0.09	417.00	-3.13	0.00 **
3-back:Trial2	-0.28	0.09	417.00	-3.10	0.00 **
1-back:Trial3	-0.13	0.09	417.20	-1.43	0.16
2-back:Trial3	-0.34	0.09	417.20	-3.71	0.00 ***
3-back:Trial3	-0.31	0.09	417.20	-3.40	0.00 ***

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Table 14

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on deoxy-Hb Measurement in Channel 16

	b	*SE*	*df*	*t*	*p*
(Intercept)	-0.12	0.06	442.00	-2.17	0.03 *
1-back	0.03	0.07	505.10	0.43	0.67
2-back	0.02	0.07	505.10	0.21	0.84
3-back	0.15	0.07	505.10	2.03	0.04 *
Trial2	0.26	0.07	505.10	3.57	0.00 ***
Trial3	0.28	0.07	505.10	3.87	0.00 ***
1-back:Trial2	-0.16	0.10	505.10	-1.51	0.13
2-back:Trial2	-0.15	0.10	505.10	-1.41	0.16
3-back:Trial2	-0.29	0.10	505.20	-2.76	0.01 **
1-back:Trial3	-0.19	0.10	505.10	-1.81	0.07
2-back:Trial3	-0.14	0.10	505.10	-1.32	0.19
3-back:Trial3	-0.37	0.10	505.10	-3.57	0.00 ***

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Apart from the abovementioned channels; in Channel 1, Channel 2 and Channel 9 random slope model lead significant improvement of the model.

Results of the random slope model revealed significant intercept in Channel 1, (b = -0.14, t(182.60) = -2.70, p < .01). The negative coefficient of the 0-back condition in trial 1 indicated significant decrease in deoxy-Hb measurements in Channel 1. For the n-back conditions and trials, deoxy-Hb measurements significantly differed in 3-back condition, (b = 0.16, t(293.10) = 2.25, p < .05) and in trial 2 and trial 3 (b = 0.25, t(409.00) = 3.65, p < .001; b = 0.29, t(409.00) = 4.15, p < .001, respectively) compared to 0-back condition in trial 1.The positive coefficients reflected significant increases in deoxy-Hb measures. Moreover, the interaction between 2-back condition in trial 3 (b = -0.25, t(409.00) = -2.61, p < .01) and the interaction between 3-back condition in trial 2 and trial 3 were statistically significant, when compared with the 0-back condition in trial 1 (b = -0.22, t(409.00) = -2.27, p < .05; b = -0.23, t(409.00) = -2.36, p < .05). The negative coefficients of the interactions revealed statistically significant decreases in deoxy-Hb measurements compared with the 0-back condition in trial 1 (Table 15).

In Channel 2, results of the random slope model revealed statistically significant intercept (b = -0.18, t(195.40) = -3. 02, p < .01). The negative coefficient of 0-back condition indicated significant decrease in deoxy-Hb measurements in Channel 2. For the n-back conditions and trial, 3-back condition (b = 0.21, t(163.90) = 2.27, p < .05), trial 2 (b = 0.27, t(453.20) = 3.50, p < .001) and trial 3 (b = 0.28, t(453.00) = 3.58, p < .001) significantly differed from the 0-back condition in trial 1. The positive coefficients reflected significant increases in deoxy-Hb measurements. Furthermore, the interaction between 1-back condition in trial 2 and trial 3, (b = -0.27, t(451.90) = -2.40, p < .05; b = -0.22, t(453.50) = -2.03, p < .05, respectively), the interaction between 2-back condition in trial 2 and trail 3 (b = -0.24, t(452.30) = -2.15, p < .05; b = -0.36, t(452.80) = -3.30, p < .01, respectively), and the interaction between 3-back condition in trial 2 and trial 3 (b = -0.25, t(451.90) = -2.24, p < .05; b = -0.31, t(453.60) = -2.81, p < .01, respectively) significantly differed from the 0-back condition in trial 1. The negative coefficients of the interactions indicated statistically significant decreases in deoxy-Hb measures compared with the 0-back condition in trial 1 (Table 16).

Table 15

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on deoxy-Hb Measurement in Channel 1

	b	*SE*	*df*	*t*	*p*
(Intercept)	-0.14	0.05	182.60	-2.70	0.01 **
1-back	0.03	0.07	223.10	0.46	0.65
2-back	0.11	0.08	124.80	1.30	0.20
3-back	0.16	0.07	293.10	2.25	0.03 *
Trial2	0.25	0.07	409.00	3.65	0.00 ***
Trial3	0.29	0.07	409.00	4.15	0.00 ***
1-back:Trial2	-0.12	0.10	409.00	-1.20	0.23
2-back:Trial2	-0.14	0.10	409.00	-1.42	0.16
3-back:Trial2	-0.22	0.10	409.00	-2.28	0.02 *
1-back:Trial3	-0.12	0.10	409.00	-1.24	0.21
2-back:Trial3	-0.25	0.10	409.00	-2.61	0.01 **
3-back:Trial3	-0.23	0.10	409.00	-2.36	0.02 *

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Table 16

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on deoxy-Hb Measurement in Channel 2

	b	*SE*	*df*	*t*	*p*
(Intercept)	-0.18	0.06	195.40	-3.02	0.00 **
1-back	0.10	0.09	219.10	1.14	0.25
2-back	0.12	0.08	237.40	1.42	0.16
3-back	0.21	0.09	163.90	2.27	0.02 *
Trial2	0.27	0.08	453.20	3.50	0.00 ***
Trial3	0.28	0.08	454.00	3.58	0.00 ***
1-back:Trial2	-0.27	0.11	451.90	-2.40	0.02 *
2-back:Trial2	-0.24	0.11	452.30	-2.15	0.03 *
3-back:Trial2	-0.25	0.11	451.90	-2.24	0.03 *
1-back:Trial3	-0.22	0.11	453.50	-2.03	0.04 *
2-back:Trial3	-0.36	0.11	452.80	-3.30	0.00 **
3-back:Trial3	-0.31	0.11	453.60	-2.81	0.01 **

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

In Channel 9, the results of the random slope model indicated that compared with the 0-back condition in trial 1, only trial 2 and trial 3 on deoxy-Hb measurements significantly varied ($b = 0.21$, $t(466.50) = 3.23$, $p < .01$; $b = 0.19$, $t(469.00) = 2.89$, $p < .01$, respectively) The positive coefficients reflected significant increases in deoxy-Hb measurements compared with the 0-back condition in trial 1 (Table 17).

Deoxy-Hb activity of the prefrontal cortex was visualized by averaging the deoxy-Hb measurements based on three trials and four n-back conditions (Figure 12).

3.3.3 Total-Hb Measurements

For the total-Hb measures, whether adding n-back conditions, trials and their interactions as fixed effects significantly improve the random intercept model was investigated. Although significant improvement of the model was found in Channel 2 (χ^2 (11) = 23.16, $p < .05$) and Channel 6 (χ^2 (11) = 22.86, $p < .05$), after FDR correction was applied to all 16-channels, statistical significance of Channel 2 and Channel 6 were lost. Therefore, none of the channels were found statistically significant. Total-Hb activity of the prefrontal cortex was visualized by averaging the total-Hb measurements based on three trials and four n-back conditions (Figure 13).

3.3.4 Oxygenation Change Measurements

For the oxygenation measures, whether adding n-back conditions, trials and their interactions as fixed effects significantly improved the random intercept model was investigated.

It was found that in Channel 1, AIC decreased from 1123.30 to 1112.80 and the log likelihood ratio test indicated statistically significance (χ^2 (11) = 32.49, $p < .05$); in Channel 2, decreased from 1354.50 to 1327.10 and the log likelihood ratio test was significant (χ^2 (11) = 49.39, $p < .05$); in Channel 4, declined from 1303.50 to 1290.70 and the log likelihood ratio test was also significant (χ^2 (11) = 34.73, $p < .05$); in Channel 6, the decrease was from 1180.20 to 1175.20 and the log likelihood ratio test showed statistically significance (χ^2 (11) = 27.02, $p < .05$); in Channel 8, it was decreased from 1211.10 to 1209.50 and the log likelihood ratio test was significant (χ^2 (11) = 23.59, $p < .05$); in Channel 12, decreased from 1245.20 to 122.80 and the log likelihood ratio test was significant (χ^2 (11) = 22.34, $p < .05$);

Table 17

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on deoxy-Hb Measurement in Channel 9

	b	*SE*	*df*	*t*	*p*
(Intercept)	-0.05	0.05	195.30	-0.97	0.33
1-back	0.02	0.07	327.70	0.31	0.75
2-back	0.08	0.08	144.80	0.98	0.33
3-back	0.08	0.07	236.00	1.04	0.30
Trial2	0.21	0.07	466.50	3.23	0.00 **
Trial3	0.19	0.07	469.00	2.89	0.00 **
1-back:Trial2	-0.05	0.09	466.50	-0.58	0.56
2-back:Trial2	-0.18	0.09	466.50	-1.94	0.05
3-back:Trial2	-0.08	0.09	466.50	-0.83	0.41
1-back:Trial3	-0.07	0.09	467.80	-0.73	0.46
2-back:Trial3	-0.15	0.09	467.80	-1.61	0.11
3-back:Trial3	-0.16	0.09	467.80	-1.72	0.09

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

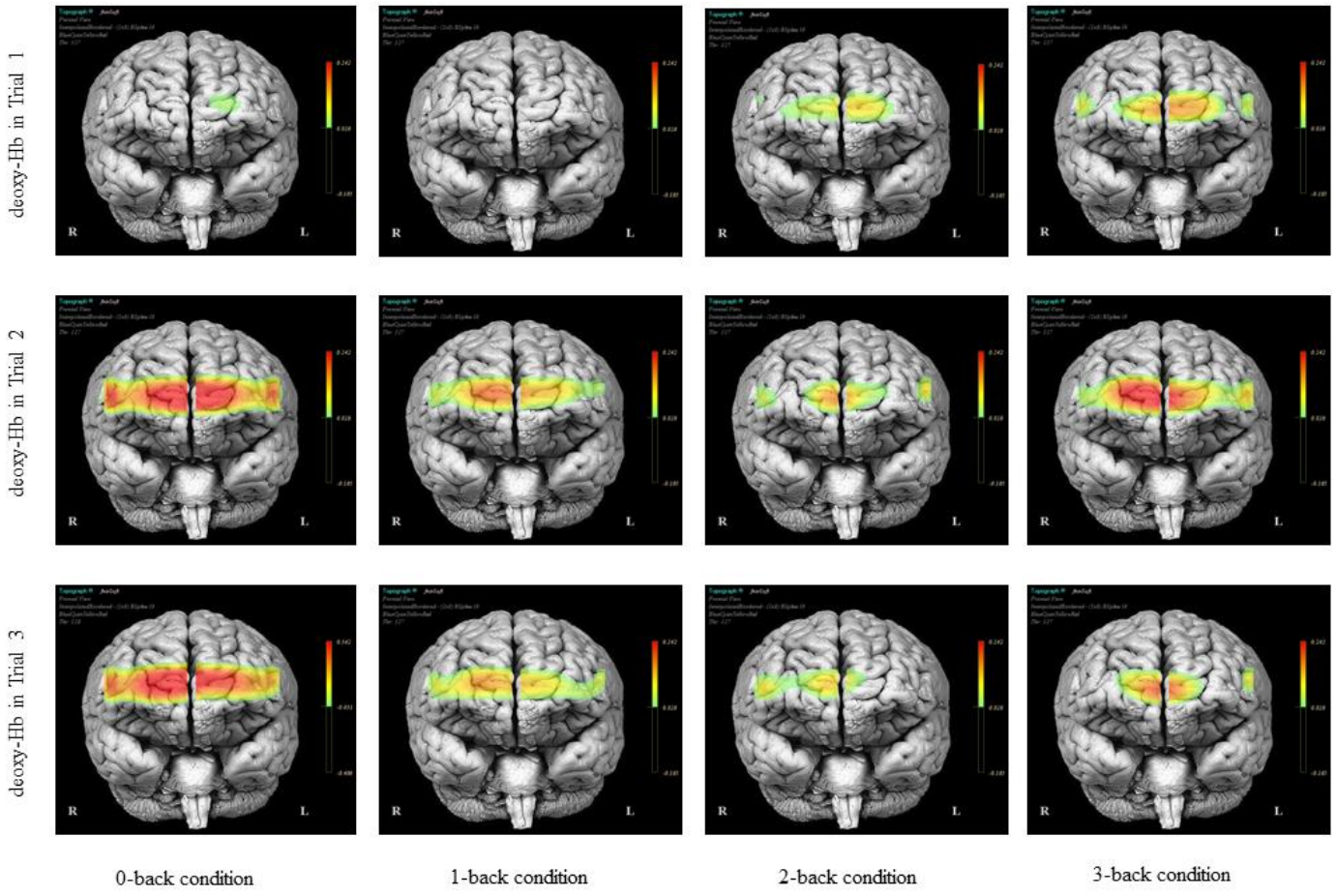

Figure 12. Mean deoxy-Hb measurements based on three trials and four n-back conditions.

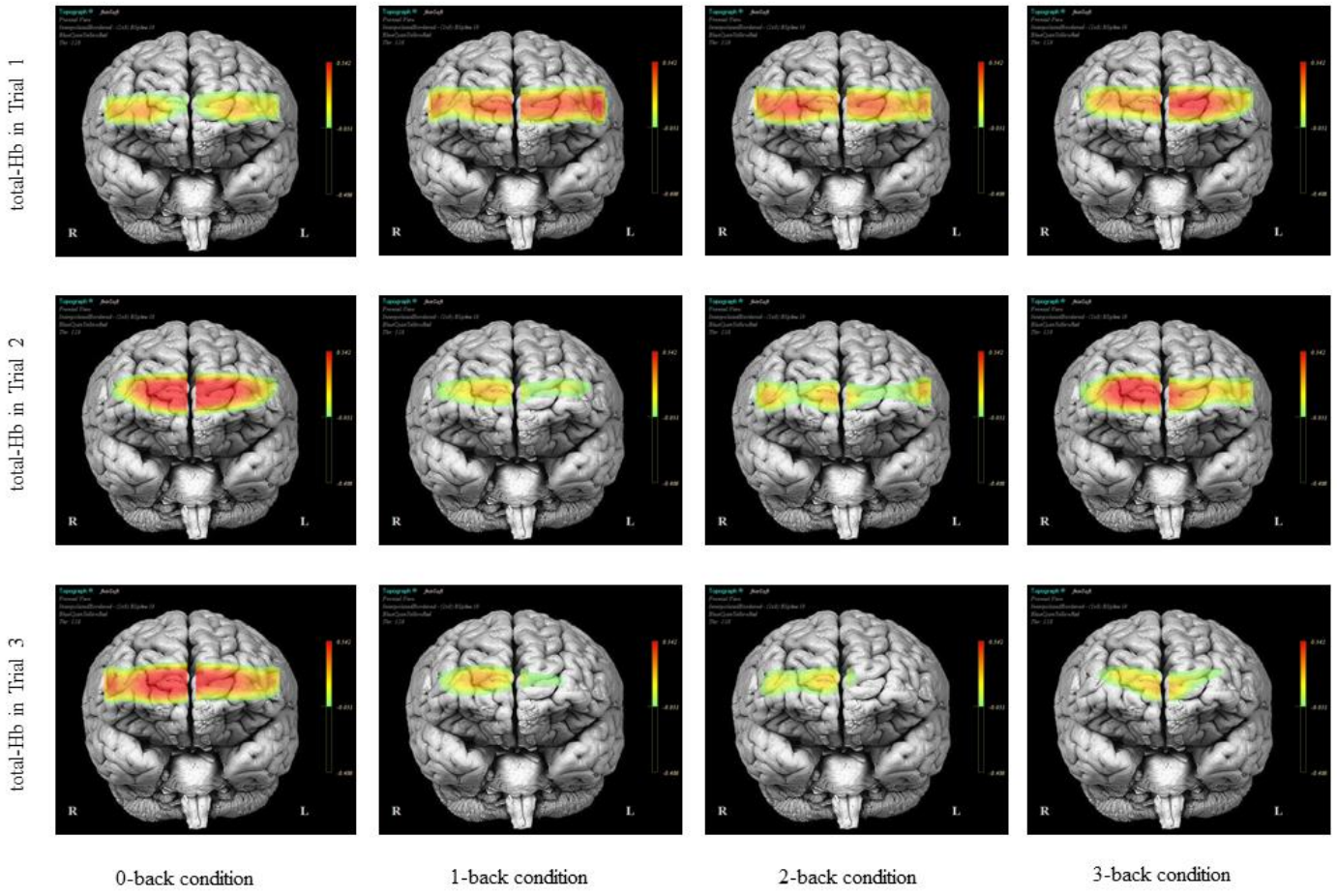

Figure 13. Mean total-Hb measurements based on three trials and four n-back conditions.

in Channel 14, declined from 1332.70 to 1323.80 and the log likelihood ratio test showed statistically significance (χ^2 (11) = 30.87, $p < .05$); in Channel 15, it was from 992.40 to 983.00 and the log likelihood ratio test was again significant (χ^2 (11) = 31.39, $p < .05$) and in Channel 16, decreased from 1394.10 to 1376.00 and the log likelihood ratio test was statistically significant (χ^2 (11) = 40.15, $p < .05$). Those findings demonstrated that including fixed effects of n-back conditions, trials and their interactions resulted in significant improvement in random intercept models for these channels.

However, adding n-back conditions, trials and their interactions as fixed effects did not significantly improve the model in Channel 3 (χ^2 (11) = 13.89, $p > .05$), Channel 5 (χ^2 (11) = 15.11, $p > .05$), Channel 7 (χ^2 (11) = 14.08, $p > .05$), Channel 9 (χ^2 (11) = 12.55, $p > .05$), Channel 10 (χ^2 (11) = 19.20, $p > .05$), Channel 11 (χ^2 (11) = 15.05, $p > .05$), Channel 13 (χ^2 (11) = 14.55, $p > .05$). Hence, other models were not tested across those channels.

Due to their statistical significance in the first model comparison, random slope model was tested in Channel 1, Channel 2, Channel 4, Channel 6, Channel 8, Channel 12, Channel 14, Channel 15, and Channel 16. Results indicated that in Channel 4, AIC decreased from 1303.50 to 1286.60 and the log likelihood ratio test was statistically significant (χ^2 (9) = 22.16, $p < .05$); in Channel 6, the decrease was from 1180.20 to 1175.90 and the log likelihood ratio test was significant (χ^2 (9) = 17.30, $p < .05$); in Channel 12, it was from 1245.20 to 1239.70 and the log likelihood ratio test was significant (χ^2 (9) = 23.20, $p < .05$); in Channel 14, declined from 1332.70 to 1380.30 and the log likelihood ratio test was again significant (χ^2 (9) = 33.53, $p < .05$); in Channel 16, the decrease was from 1394.10 to 1376.20 and the log likelihood ratio test was statistically significant (χ^2 (9) = 17.77, $p < .05$) These results revealed random slope model also significantly improved the model.

On the other hand, the random slope model did not improve the model in following channels: in Channel 1 (χ^2 (9) = 6.55, $p > .05$); Channel 2 (χ^2 (9) = 10.39, $p > .05$); Channel 8 (χ^2 (9) = 14.27, $p > .05$) and Channel 15 (χ^2 (9) = 10.66, $p > .05$). Therefore, random intercept model was accepted as the best fitted model for those channels.

After FDR correction was applied to all 16-channels, Channel 1, 2, 4, 8, 12, 14, and 15 were found statistically significant.

As it was pointed out above, only the random intercept model improved the empty model in Channel 1, Channel 2, Channel 8, and Channel 15. In Channel 1, results of the random intercept model indicated that only in trail 2 and trial 3 oxygenation measurements significantly differed from the 0-back condition in trial 1 (b = -0.48, t(450.10) = -2.96, $p < .05$; b = -0.41, t(450.10) = -2.55, $p < .05$). The negative coefficient of trial 2 and trial 3 showed significant decrease in oxygenation measurements, compared with the 0-back condition in trial 1 (Table 18).

In Channel 2, results of the random intercept model demonstrated statistically significant intercept (b = 0.35, t(540.20) = 2.94, $p < .01$). In addition, compared with the 0-back condition in trial 1, 3-back condition (b = -0.56, t(494.50) = -3.33, $p < .001$), trial 2 and trial 3 (b = -0.78, t(495.40) = -4.61, $p < .001$; b = -0.65, t(495.50) = -3.87, $p < .001$, respectively) significantly differed. The negative coefficients showed significant decreases in oxygenation measurements. The interactions between 3-back condition in trial 2 and trial 3 (b = 0.81, t(495.00) = 3.38, $p < .001$; b = 0.53, t(494.50) = 2.24, $p < .05$, respectively) and the interaction between 2-back condition in trial 2 were significantly varied from 0-back condition in trial 1 (b = 0.59, t(496.80) = 2.45, $p < .05$). The positive coefficients of interactions showed significant increases in oxygenation measurements when compared with 0-back condition in trial 1 (Table 19).

In Channel 8, results of the random intercept model indicated significant intercept (b = -0.29, t(514.30) = -2.02, $p < .05$). The negative coefficient of trial 2 showed significant decrease in oxygenation measurements compared with 0-back condition in trial 1 (Table 20).

In Channel 15, results of the random intercept model revealed that compared to 0-back condition in trial 1, 3-back condition significantly differed on oxygenation measurements (b = -0.39, t(417.10) = -2.48, $p < .05$), showing significant decrease in measurements. The interactions between 3-back condition in trial 2, significantly differed from 0-back condition in trial 1 (b = 0.67, t(417.10) = 3.03, $p < .0$). The positive coefficient of the interaction demonstrated statistically significant increase in oxygenation measurements (Table 21).

Table 18

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on Oxygenation Change Measurements in Channel 1

	b	*SE*	*df*	*t*	*p*
(Intercept)	0.21	0.11	490.40	1.83	0.07
1-back	0.12	0.16	450.90	0.73	0.47
2-back	-0.09	0.16	450.10	-0.56	0.58
3-back	-0.24	0.16	450.10	-1.49	0.14
Trial2	-0.48	0.16	450.10	-2.96	0.00 **
Trial3	-0.41	0.16	450.10	-2.54	0.01 *
1-back:Trial2	-0.07	0.23	450.50	-0.30	0.77
2-back:Trial2	0.25	0.23	450.10	1.11	0.27
3-back:Trial2	0.42	0.23	450.10	1.87	0.06
1-back:Trial3	-0.13	0.23	450.50	-0.59	0.56
2-back:Trial3	0.14	0.23	450.10	0.62	0.53
3-back:Trial3	0.12	0.23	450.10	0.52	0.60

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Table 19

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on Oxygenation Change Measurements in Channel 2

	b	*SE*	*df*	*t*	*p*
(Intercept)	0.35	0.12	540.20	2.94	0.00 **
1-back	-0.02	0.17	495.40	-0.10	0.92
2-back	-0.20	0.17	495.40	-1.20	0.23
3-back	-0.56	0.17	494.50	-3.33	0.00 ***
Trial2	-0.78	0.17	495.40	-4.61	0.00 ***
Trial3	-0.65	0.17	494.50	-3.87	0.00 ***
1-back:Trial2	0.12	0.24	495.00	0.49	0.62
2-back:Trial2	0.59	0.24	496.80	2.48	0.01 *
3-back:Trial2	0.81	0.24	495.00	3.38	0.00 ***
1-back:Trial3	0.07	0.24	495.00	0.31	0.76
2-back:Trial3	0.40	0.24	495.00	1.68	0.09
3-back:Trial3	0.53	0.24	494.50	2.24	0.03 *

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Table 20

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on Oxygenation Change Measurements in Channel 8

	b	*SE*	*df*	*t*	*p*
(Intercept)	0.06	0.10	559.70	0.58	0.56
1-back	0.16	0.14	514.30	1.15	0.25
2-back	-0.04	0.14	514.30	-0.29	0.77
3-back	-0.11	0.14	514.30	-0.79	0.43
Trial2	-0.29	0.14	514.30	-2.02	0.04 *
Trial3	-0.14	0.14	514.30	-0.96	0.34
1-back:Trial2	-0.19	0.20	514.30	-0.95	0.35
2-back:Trial2	0.14	0.20	514.30	0.68	0.50
3-back:Trial2	0.15	0.20	514.30	0.74	0.46
1-back:Trial3	-0.36	0.20	515.10	-1.78	0.08
2-back:Trial3	-0.03	0.20	515.10	-0.17	0.87
3-back:Trial3	-0.04	0.20	514.70	-0.19	0.85

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Table 21

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on Oxygenation Change Measurements in Channel 15

	b	*SE*	*df*	*t*	*p*
(Intercept)	0.19	0.11	170.10	1.67	0.10
1-back	0.13	0.15	233.30	0.82	0.42
2-back	-0.13	0.16	189.10	-0.82	0.41
3-back	-0.39	0.16	215.50	-2.45	0.02 *
Trial2	-0.54	0.15	379.40	-3.63	0.00 ***
Trial3	-0.36	0.15	381.90	-2.39	0.02 *
1-back:Trial2	0.09	0.21	379.40	0.45	0.66
2-back:Trial2	0.42	0.21	379.40	2.01	0.05 *
3-back:Trial2	0.67	0.21	379.40	3.17	0.00 **
1-back:Trial3	-0.15	0.21	380.70	-0.68	0.49
2-back:Trial3	0.14	0.21	380.70	0.67	0.51
3-back:Trial3	0.30	0.21	380.70	1.40	0.16

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Other than the aforementioned channels, in Channel 4, Channel 12 and Channel 14 random slope model resulted in significant improvement compared to random intercept model.

In Channel 4, results of the random slope model revealed significant intercept (b = 0.28, t(233.50) = 2.59, p < .05). The positive coefficient of 0-back condition in trial 1 indicated significant increase in oxygenation measurements. In addition, compared with the 0-back condition in trial 1, oxygenation measurements significantly varied in trial 2 and trial 3 (b = -0.49, t(423.00) = -3.45, p < .001; b = -0.44, t(423.00) = -3.12, p < .01, respectively). The negative coefficients reflected significant decreases in oxygenation measurements when compared with the 0-back condition in trial 1 (Table 22).

In Channel 12, results of the random slope model indicated oxygenation measurements significantly differed in trial 2 from 0-back condition in trial 1 (b = -0.37, t(469.10) = -2.70, p < .01). The negative coefficient showed significant decrease in oxygenation measurements compared with 0-back condition in trial 1 (Table 23).

In Channel 14, results of the random slope model demonstrated that compared with the 0-back condition in trial 1, oxygenation measurements significantly differed in trial 2 and trial 3 (b = -0.56, t(422.50) = -3.89, p < .001; b = -0.32, t(422.50) = -2.22, p < .05, respectively). The negative coefficients showed significant decreases in oxygenation measurements compared with 0-back condition in trial 1. Also, the interaction between 3-back in trial 2 significantly differed from 0-back condition in trial 1 (b = 0.55, t(422.10) = 2.68, p < .01). The positive coefficient of the interaction indicated significant increase in oxygenation measurements when compared with the 0-back condition in trial 1 (Table 23). Oxygenation change of the prefrontal cortex was depicted by averaging the total-Hb measurements based on three trials and four n-back conditions (Figure 14).

Table 22

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on Oxygenation Change Measurements in Channel 4

	b	*SE*	*df*	*t*	*p*
(Intercept)	0.28	0.11	233.50	2.59	0.01 *
1-back	0.02	0.15	296.00	0.14	0.89
2-back	-0.10	0.15	217.20	-0.63	0.53
3-back	-0.24	0.17	158.90	-1.48	0.14
Trial2	-0.48	0.14	423.00	-3.45	0.00 ***
Trial3	-0.44	0.14	423.00	-3.12	0.00 **
1-back:Trial2	0.08	0.20	423.00	0.38	0.70
2-back:Trial2	0.32	0.20	423.00	1.60	0.11
3-back:Trial2	0.34	0.20	423.00	1.71	0.09
1-back:Trial3	-0.09	0.20	423.00	-0.46	0.65
2-back:Trial3	0.24	0.20	423.00	1.21	0.23
3-back:Trial3	0.15	0.20	423.00	0.76	0.45

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Table 23

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on Oxygenation Change Measurements in Channel 12

	b	*SE*	*df*	*t*	*p*
(Intercept)	0.19	0.10	297.60	1.84	0.07
1-back	0.04	0.14	413.50	0.27	0.79
2-back	-0.04	0.15	209.50	-0.24	0.81
3-back	-0.22	0.16	181.00	-1.39	0.17
Trial2	-0.37	0.14	469.10	-2.70	0.01 **
Trial3	-0.24	0.14	469.80	-1.72	0.09
1-back:Trial2	-0.05	0.19	469.10	-0.26	0.79
2-back:Trial2	0.21	0.19	469.10	1.11	0.27
3-back:Trial2	0.34	0.19	469.10	1.76	0.08
1-back:Trial3	-0.16	0.19	469.50	-0.82	0.42
2-back:Trial3	0.02	0.19	469.50	0.12	0.91
3-back:Trial3	0.08	0.19	469.50	0.43	0.67

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Table 24

Multilevel Analyses of the Effect of N-back Condition, Trial and Their Interaction on Oxygenation Change Measurements in Channel 14

	b	*SE*	*df*	*t*	*p*
(Intercept)	0.21	0.11	237.10	1.94	0.05
1-back	0.07	0.15	284.90	0.48	0.63
2-back	-0.02	0.16	187.70	-0.13	0.89
3-back	-0.23	0.17	153.80	-1.34	0.18
Trial2	-0.56	0.14	422.50	-3.89	0.001 ***
Trial3	-0.32	0.14	422.50	-2.22	0.03 *
1-back:Trial2	0.17	0.20	423.20	0.82	0.41
2-back:Trial2	0.36	0.20	421.70	1.77	0.08
3-back:Trial2	0.55	0.20	422.10	2.68	0.00 **
1-back:Trial3	-0.21	0.20	423.20	-1.01	0.31
2-back:Trial3	0.01	0.20	421.70	0.05	0.96
3-back:Trial3	0.16	0.20	421.70	0.78	0.44

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

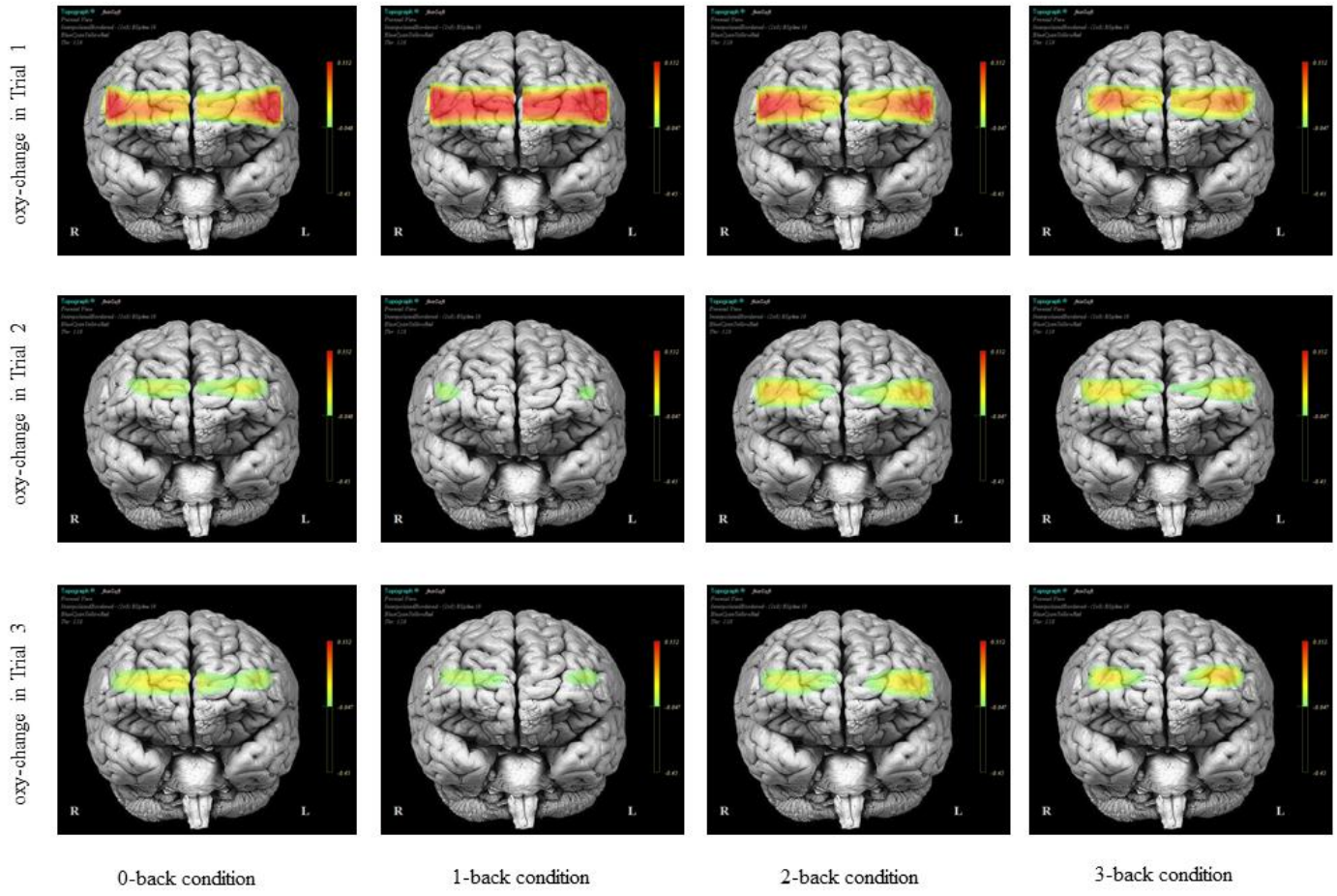

Figure 14. Mean oxy-change measurements based on three trials and four n-back conditions.

3.4 Validity of Behavioral and Hemodynamic Data

For both behavioral and hemodynamic measurements, whether adding gender as a fixed effect improves the model was investigated. If adding gender improves the model, then this indicates that gender of the participants has an effect on measurements; but if not then this finding was accepted as a proof to construct validity of both behavioral and hemodynamic measurements.

3.4.1 Validity of Behavioral Data

For the number of detected target letters, number of detected non-target letters, the reaction time during target letter detection, and the reaction time during non-target letter detection, whether adding gender as a fixed effect to the random slope models, which were found significant previously, were investigated.

The number of detected target letters, the number of detected non-target letters, and the reaction time during non-target letter detection results indicated that the random slopes model did not improve the models (χ^2 (12) = 13.39, $p > .05$; χ^2 (12) = 13.39, $p > .05$; χ^2 (12) = 17.70, $p > .05$, respectively). These results revealed that including fixed effect of gender to the random slope models did not result in differences in the number of detected target letters, the number of detected non-target letters results, and the reaction time during non-target letter detection. Therefore, it can be concluded that similar results of the number of detected target letters, the number of detected non-target letters results, and the reaction time during non-target letter detection with respect to gender-differences would be a proof for the construct validity of measures.

On the other hand, for the reaction time during target letter detection, results indicated that the random slopes model with the fixed effect of gender improved the model (χ^2 (12) = 45.95, $p < .05$). This result showed that including gender to the random slope model lead to differences in the reaction time during target letter detection. For the reaction time during target letter detection, AIC decreased from -45.32 to -432.36 and BIC declined from -32.31 to -280.63 and the log likelihood ratio test indicated statistically significance (χ^2 (12) = 45.95, $p < .05$). Results of the random slope model indicated significant intercept, $b = 0.50$, $t(136.20) = 13.49$, $p < .05$. The reaction time during target letter detection significantly differed from 1-back condition, $b = 0.11$, $t(210.00) = 2.27$, $p < .05$; 2-back condition, $b = 0.20$,

$t(139.60) = 3.93$, $p < .05$; and 3-back condition, $b = 0.35$, $t(101.10) = 5.88$, $p < .05$, compared to 0-back condition of males in trial 1. The positive coefficients reflected that the reaction time during target letter detection significantly increased in 1-, 2-, and 3-back conditions compared with the 0-back condition in trial 1 of males. In addition, the interaction of 2-back in trial 2 of females and 3-back condition in trial 2 of females was significantly differed from 0-back condition in trial 1 of males ($b = -0.19$, $t(423.00) = -2.52$, $p < .05$; $b = -0.23$, $t(423.00) = -2.52$, $p < .05$). The negative coefficients indicated that the reaction time during target letter detection significantly decreased in 2- and 3-back conditions compared with the 0-back condition in trial 1 of males (Table 25).

3.4.2 Validity of Hemodynamic Data

For the oxy-Hb measures, whether adding gender to the random intercept model improves the model was investigated. This model was tested only in Channel 2; because it was the only channel that remained significant after FDR correction. It was found that the random intercept model (with the fixed effect of gender) did not improve the model (χ^2 (12) = 20.41, $p > .05$). Therefore, oxy-Hb measure did not significantly differ for males and females.

For the deoxy-Hb measures, gender was added to the random intercept model in Channel 4, 6, 8, 11, 12, 14, 15 and 16, because random intercept model was found significant previously. The random intercepts with the fixed effect of gender did not improve the model (in Channel 4, χ^2 (12) = 3.34, $p > .05$; Channel 6, χ^2 (12) = 11.60, $p > .05$; Channel 8, χ^2 (12) = 11.85, $p > .05$; Channel 11, χ^2 (12) = 16.11, $p > .05$; Channel 12, χ^2 (12) = 9.62, $p > .05$; Channel 15, χ^2 (12) = 11.90, $p < .05$; Channel 16, χ^2 (12) = 6.54, $p < .05$; except in Channel 14 (χ^2 (12) = 23.09, $p < .05$).

In Channel 14, the log likelihood ratio test indicated statistical significance (χ^2 (12) = 23.09, $p < .05$), but the AIC remained same. Results of the random intercept model with the fixed effect of gender indicated, 1-back condition, $b = -0.29$, $t(514.20) = -3.22$, $p < .05$; significantly differed from 0-back condition of males in trial 1. The negative coefficient signified that deoxy-Hb measurements decreased in 1-back condition compared with the 0-back condition in trial 1 of males. Additionally, the interaction between 1-back condition of females significantly varied from 0-back condition of males in trial 1 ($b = 0.51$, $t(514.20) = 3.49$, $p < .05$).

Table 25

Multilevel Analyses of the Effects of N-back Condition, Trial, Gender, and Their Interactions on Reaction Time During Target Letter Detection

	b	*SE*	*df*	*t*	*p*
(Intercept)	0.50	0.04	136.20	13.49	2e-16 ***
1-back	0.11	0.05	210.00	2.27	0.02 *
2-back	0.20	0.05	139.60	3.93	0.000 ***
3-back	0.35	0.06	101.10	5.88	5.46e-08 ***
Trial2	0.03	0.04	423.00	0.73	0.47
Trial3	0.02	0.04	423.00	0.51	0.61
Female	0.02	0.05	136.20	0.40	0.69
1-back:Trial2	-0.07	0.06	423.00	-1.09	0.28
2-back:Trial2	0.06	0.06	423.00	-1.04	0.30
3-back:Trial2	0.08	0.06	423.00	1.30	0.19
1-back:Trial3	-0.01	0.06	423.00	-0.24	0.81
2-back:Trial3	-0.02	0.06	423.00	-0.42	0.68
3-back:Trial3	-0.05	0.06	423.00	-0.92	0.36
1-back:Female	0.06	0.06	210.00	1.02	0.31
2-back:Female	0.06	0.07	139.60	0.90	0.37
3-back:Female	0.09	0.08	101.10	1.19	0.24
Trial2:Female	-0.03	0.05	423.00	-0.49	0.62

Table 25 (continued)

Trial3:Female	-0.04	0.05	423.00	-0.81	0.42
1-back:Trial2:Female	0.01	0.08	423.00	0.11	0.92
2-back:Trial2:Female	-0.19	0.08	423.00	-2.52	0.01 *
3-back:Trial2:Female	-0.23	0.08	423.00	-3.09	0.00 **
1-back:Trial3:Female	-0.06	0.08	423.00	-0.79	0.43
2-back:Trial3: Female	-0.02	0.08	423.00	-0.26	0.79
3-back:Trial:Female	-0.07	0.08	423.00	-0.90	0.37

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Also, the interaction between 1-back condition in trial 2 of females and the interaction between 1-back condition in trial 3 of females were significantly differed from 0-back condition in trial 1 of males (b = -0.51, t(514.20) = -2.51, p < .05; b = -0.45, t(514.20) = -2.21, p < .05; respectively) (Table 26).

Table 26

Multilevel Analyses of the Effects of N-back Condition, Trial, Gender, and Their Interactions on deoxy-Hb Measurements in Channel 14

	b	*SE*	*df*	*t*	*p*
(Intercept)	-0.03	0.07	551.50	-0.48	0.63
1-back	-0.29	0.09	514.20	-3.22	0.00 **
2-back	0.02	0.09	515.30	0.23	0.82
3-back	0.06	0.09	515.30	0.61	0.54
Trial2	0.14	0.09	515.30	1.15	0.13
Trial3	0.13	0.09	515.30	1.43	0.15
Female	-0.04	0.11	550.80	-0.40	0.69
1-back:Trial2	0.18	0.13	514.20	1.41	0.16
2-back:Trial2	-0.06	0.13	514.80	-0.48	0.63
3-back:Trial2	-0.15	0.13	514.80	-1.16	0.25
1-back:Trial3	0.22	0.13	514.20	1.70	0.09
2-back:Trial3	-0.06	0.13	514.80	-0.49	0.62
3-back:Trial3	-0.21	0.13	514.80	-1.68	0.09
1-back:Female	0.51	0.15	514.20	3.49	0.00 ***
2-back:Female	0.06	0.14	514.60	0.40	0.69
3-back:Female	-0.04	0.14	514.60	-0.30	0.77
Trial2:Female	0.07	0.14	514.60	0.50	0.61

Table 26 (continued)

Trial3:Female	0.09	0.14	514.60	0.63	0.53
1-back:Trial2:Female	-0.51	0.20	514.20	-2.51	0.01 *
2-back:Trial2:Female	-0.20	0.20	514.40	-0.99	0.32
3-back:Trial2:Female	0.06	0.20	514.90	0.29	0.77
1-back:Trial3:Female	-0.45	0.20	514.20	-2.21	0.03 *
2-back:Trial3: Female	-0.12	0.20	514.40	-0.59	0.55
3-back:Trial:Female	0.02	0.20	514.40	0.14	0.89

Note. The lmerTest package was used to compute approximate *p*-values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed).

Also, gender was added to random slope models in Channel 1, 2, and 9, because of their significant improvement in random slope model. Results indicated random slope with the fixed effect of gender did not significantly improve the model in Channel 1, χ^2 (12) = 14.73, $p > .05$; Channel 2, χ^2 (12) = 14.02, $p > .05$; Channel 9, χ^2 (12) = 12.71.

Due to none of the channels improve any models in total-Hb measures; the effect of gender was not investigated in total-Hb measures.

For the oxygenation measures, gender was added to the random intercept model in Channel 1, 2, 8, and 15, due to their previous improvement in this model. It was found that the random intercept model with the fixed effect of gender did not improve the model fit in any of the channels (in Channel 1, χ^2 (12) = 16.19, $p > .05$; Channel 2, χ^2 (12) = 16.73, $p > .05$; Channel 8, χ^2 (12) = 12.72, $p > .05$; Channel 15, χ^2 (12) = 16.53, $p > .05$). Gender was also added to random slope models in Channel 1, 2, and 9, due to their improvement in random slope model. Results indicated that random slope model with the fixed effect of gender did not improve the model in Channel 4, χ^2 (12) = 11.71, $p > .05$; Channel 12, χ^2 (12) = 8.50, $p > .05$; and Channel 14, χ^2 (12) = 11.07.

Overall construct validity examination indicated that males and females did not differ both in behavioral measures (except the reaction time during target letter detection) hemodynamic measures (except Channel 14).

3.5 Reliability of Behavioral and Hemodynamic Data

For estimating the stability of behavioral and hemodynamic measures across first and second sessions' test-retest reliability, for the reliability of behavioral and hemodynamic measures across two different versions of the same task, alternate forms reliability results were examined by using the parameters of ICCs.

As it was stated before, a unified ANOVA model is the statistical foundation of ICCs. In this thesis, for the reliability of single measurements a two-way random effects model, which is referred to as ICC(2,1), in order to determine the reliability in of terms of absolute agreement and two-way mixed effects model, which is denoted as ICC(3,1), to determine the reliability in of terms of consistency were used in both behavioral and hemodynamic measurements of test-retest and alternate forms reliability investigations.

3.5.1 Test-retest Reliability of Behavioral Data

For the number of detected target letters, good agreement and consistency were found between two sessions (ICC(2,1) = 0.60; ICC(3,1) = 0.74, respectively). On the other hand, poor agreement and fair consistency were found for the number of detected non-target across two sessions (ICC(2,1) = 0.37; ICC(3,1) = 0.51, respectively). For the reaction time during target letter detection, good agreement and excellent consistency were found between two sessions (ICC(2,1) = 0.68; ICC(3,1) = 0.83, respectively). The reaction time during non-target letter detection revealed fair agreement and consistency across two sessions (ICC(2,1) = 0.55; ICC(3,1) = 0.54, respectively). Those findings revealed that behavioral measures were stable over time (Table 27).

3.5.2 Test-retest Reliability of Hemodynamic Data

For the oxy-Hb measures, fair agreement and consistency were found for Channel 2 (ICC(2,1) = 0.57; ICC(3,1) = 0.59), Channel 3 (ICC(2,1) = 0.50; ICC(3,1) = 0.52, Channel 4 (ICC(2,1) = 0.56; ICC(3,1) = 0.56, Channel 6 (ICC(2,1) = 0.52); ICC(3,1) = 0.51), Channel 7 (ICC(2,1) = 0.50; ICC(3,1) = 0.48), Channel 9 (ICC(2,1) = 0.51; ICC(3,1) = 0.50), and Channel 16 (ICC(2,1) = 0.59; ICC(3,1) = 0.57) between two sessions. For Channel 8 (ICC(2,1) = 0.62; ICC(3,1) = 0.60), Channel 10 (ICC(2,1) = 0.61; ICC(3,1) = 0.60), Channel 12 (ICC(2,1) = 0.62; ICC(3,1) = 0.62) and Channel 14 (ICC(2,1) = 0.69; ICC(3,1) = 0.68) good agreement and consistency were observed. Although, due to the negative reliability estimates of Channel 1, 5, 11, 13, and 15, which were accepted as not reliable, overall findings of test-retest analysis indicated stability of oxy-Hb measures over time (Table 28).

For the deoxy-Hb measures, fair agreement and consistency were found only for Channel 7 (ICC(2,1) = 0.51; ICC(3,1) = 0.57) between two sessions. Channel 1, Channel 2, Channel 5, Channel 6, Channel 8, Channel 9, Channel 10, Channel 11, and Channel 15 were not found statistically significant. In addition, because the reliability estimates of Channel 3, 4, 12, and 16 were negative those channels were accepted as not reliable. Results of the test-retest reliability analysis showed that deoxy-Hb measures were not stable across time (Table 29).

Table 27

Intraclass Correlation Coefficients with 95% Confidence Interval for Behavioral Measurements of Test-retest Reliability Study

Behavioral Measurements	ICC(2,1) (95%-CI)	ICC(3,1) (95%-CI)
Target	0.60 (-0.01, 0.87) *	0.74 (0.36, 0.91) *
Non-target	0.37 (-0.10, 0.73) *	0.51 (-0.01, 0.81) *
Reaction Time Target	0.68 (-0.01, 0.91) *	0.83 (0.54, 0.94) *
Reaction Time Non-Target	0.55 (0.06, 0.83) *	0.54 (0.04, 0.83) *

Note. Significant values were shown with (*) $p < 0.05$.

Table 28

Intraclass Correlation Coefficients with 95% Confidence Interval for oxy-Hb Measurements of Test-retest Reliability Study

Channels	ICC(2,1) (95%-CI)	ICC(3,1) (95%-CI)
Channel 1	0	0
Channel 2	0.57 (0.09, 0.84) *	0.59 (0.08, 0.85) *
Channel 3	0.50 (-0.11, 0.83) *	0.52 (-0.10, 0.82) *
Channel 4	0.56 (0.07, 0.84) *	0.56 (0.06, 0.83) *
Channel 5	0	0
Channel 6	0.52 (-0.00, 0.82) *	0.51 (-0.01, 0.81) *
Channel 7	0.50 (-0.06, 0.81) *	0.48 (-0.07, 0.81) *
Channel 8	0.62 (0.14, 0.86) *	0.60 (0.13, 0.85) *
Channel 9	0.51 (-0.02, 0.81) *	0.50 (-0.02, 0.81) *
Channel 10	0.61 (0.14, 0.86) *	0.60 (0.13, 0.85) *
Channel 11	0	0
Channel 12	0.62 (0.14, 0.87) *	0.62 (0.13, 0.87) *
Channel 13	0.34 (-0.22, 0.73)	0.33 (-0.22, 0.72)
Channel 14	0.69 (0.27, 0.89) *	0.68 (0.25, 0.88) *
Channel 15	0	0
Channel 16	0.59 (0.10, 0.85) *	0.57 (0.09, 084) *

Note. Significant values were shown with (*) $p < 0.05$.

Table 29

Intraclass Correlation Coefficients with 95% Confidence Interval for deoxy-Hb Measurements of Test-retest Reliability Study

Channels	ICC(2,1) (95%-CI)	ICC(3,1) (95%-CI)
Channel 1	0.04 (-0.74, 0.70)	0.03 (-0.65, 0.68)
Channel 2	0.06 (-0.54, 0.59)	0.06 (-0.49, 0.57)
Channel 3	0	0
Channel 4	0	0
Channel 5	0.07 (-0.34, 0.58)	0.09 (-0.54, 0.66)
Channel 6	0.12 (-0.40, 0.59)	0.12 (-0.42, 0.60)
Channel 7	0.51 (0.02, 0.81) *	0.57 (0.05, 0.85) *
Channel 8	0.24 (-0.18, 0.63)	0.30 (-0.26, 0.70)
Channel 9	0.32 (-0.19, 0.71)	0.34 (-0.22, 0.73)
Channel 10	0.07 (0.14, 0.86)	0.07 (-0.46, 0.56)
Channel 11	0.16 (-0.30, 0.60)	0.18 (-0.37, 0.64)
Channel 12	0	0
Channel 13	0	0
Channel 14	0	0
Channel 15	0.18 (-0.61, 0.74)	0.17 (-0.52, 0.72)
Channel 16	0	0

Note. Significant values were shown with (*) $p < 0.05$.

For the total-Hb measures, fair agreement and good consistency was found for Channel 2 (ICC(2,1) = 0.58; ICC(3,1) = 0.62), excellent agreement and consistency was observed for Channel 3 (ICC(2,1) = 0.80; ICC(3,1) = 0.81), good agreement and excellent consistency was identified for Channel 5 (ICC(2,1) = 0.74; ICC(3,1) = 0.75), and fair agreement and consistency was observed for Channel 14 (ICC(2,1) = 0.47; ICC(3,1) = 0.45) between two sessions. No statistical significance was found in Channel 4, Channel 6, Channel 7, Channel 8, Channel 9, Channel 10, Channel 11, Channel 12, Channel 13, Channel 15 and Channel 16. Negative reliability estimate was obtained for Channel 1, which was accepted as not reliable. Findings of the test-retest reliability analysis showed that for most of the channels, total-Hb measures were not stable across time (Table 30).

For the oxygenation measurements, fair agreement and consistency were found for Channel 4 (ICC(2,1) = 0.58; ICC(3,1) = 0.56), Channel 6 (ICC(2,1) = 0.57; ICC(3,1) = 0.58), Channel 13 (ICC(2,1) = 0.49; ICC(3,1) = 0.52), good agreement and consistency were obtained for Channel 3 (ICC(2,1) = 0.62; ICC(3,1) = 0.61), Channel 7 (ICC(2,1) = 0.72; ICC(3,1) = 0.70), Channel 9 (ICC(2,1) = 0.69; ICC(3,1) = 0.70), Channel 10 (ICC(2,1) = 0.71; ICC(3,1) = 0.74) and Channel 14 (ICC(2,1) = 0.63; ICC(3,1) = 0.62) between two sessions. For Channel 1, fair agreement and good consistency (ICC(2,1) = 0.57; ICC(3,1) = 0.67), for Channel 16 good agreement and fair consistency (ICC(2,1) = 0.61; ICC(3,1) = 0.59), for Channel 5 good agreement and excellent consistency (ICC(2,1) = 0.71; ICC(3,1) = 0.77), for Channel 8 (ICC(2,1) = 0.75; ICC(3,1) = 0.77) and Channel 12 (ICC(2,1) = 0.76; ICC(3,1) = 0.80) excellent agreement and consistency were found across two sessions. On the other hand, for Channel 1, 2, and 15 no statistically significance was obtained. Although none significant and negative values were found, overall results of the test-retest analysis revealed stability of oxygenation measurements over time (Table 31).

3.5.3 Alternate Forms Reliability of Behavioral Data

For the number of detected target and non-target letters, fair agreement and consistency were found between two forms of the same task (ICC(2,1) = 0.55; ICC(3,1) = 0.55; (ICC(2,1) = 0.42; ICC(3,1) = 0.55; respectively). For the reaction time during target letter detection, good agreement and consistency were found

Table 30

Intraclass Correlation Coefficients with 95% Confidence Interval for total-Hb Measurements of Test-retest Reliability Study

Channels	ICC(2,1) (95%-CI)	ICC(3,1) (95%-CI)
Channel 1	0	0
Channel 2	0.58 (0.11, 0.85) *	0.62 (0.14, 0.87) *
Channel 3	0.80 (0.45, 0.94) *	0.81 (0.48, 0.94) *
Channel 4	0.38 (-0.18, 0.75)	0.37 (-0.18, 0.74)
Channel 5	0.74 (0.32, 0.92) *	0.75 (0.31, 0.93) *
Channel 6	0.28 (-0.28, 0.70)	0.28 (-0.28, 0.69)
Channel 7	0.32 (-0.25, 0.72)	0.32 (-0.26, 0.73)
Channel 8	0.38 (-0.19, 0.75)	0.37 (-0.18, 0.74)
Channel 9	0.35 (-0.24, 0.74)	0.35 (-0.24, 0.74)
Channel 10	0.41 (-0.17, 0.77)	0.39 (-0.16, 0.75)
Channel 11	0.26 (-0.34, 0.69)	0.25 (-0.31, 0.67)
Channel 12	0.31 (-0.31, 0.73)	0.30 (-0.28, 0.72)
Channel 13	0.00 (-0.57, 0.54)	0.00 (-0.51, 0.52)
Channel 14	0.47 (-0.09, 0.79) *	0.45 (-0.08, 0.78) *
Channel 15	0.25 (-0.30, 0.71)	0.27 (-0.36, 0.73)
Channel 16	0.35 (-0.19, 0.73)	0.35 (-0.20, 0.73)

Note. Significant values were shown with (*) $p < 0.05$.

Table 31

Intraclass Correlation Coefficients with 95% Confidence Interval for Oxygenation Change Measurements of Test-retest Reliability Study

Channels	ICC(2,1) (95%-CI)	ICC(3,1) (95%-CI)
Channel 1	0.20 (-0.29, 0.68)	0.24 (-0.42, 0.74)
Channel 2	0.33 (-0.28, 0.74)	0.32 (-0.26, 0.73)
Channel 3	0.62 (0.10, 0.87) *	0.61 (0.08, 0.87) *
Channel 4	0.58 (0.08, 0.84) *	0.56 (0.07, 0.84) *
Channel 5	0.71 (0.19, 0.91) *	0.77 (0.34, 0.93) *
Channel 6	0.57 (0.10, 0.83) *	0.58 (0.09, 0.84) *
Channel 7	0.72 (0.29, 0.91) *	0.70 (0.27, 0.90) *
Channel 8	0.75 (0.39, 0.91) *	0.77 (0.41, 0.92) *
Channel 9	0.69 (0.29, 0.89) *	0.70 (0.29, 0.89) *
Channel 10	0.71 (0.32, 0.90) *	0.74 (0.36, 0.91) *
Channel 11	0.57 (0.04, 0.84) *	0.67 (0.24, 0.88) *
Channel 12	0.76 (0.35, 0.92) *	0.80 (0.46, 0.93) *
Channel 13	0.49 (0.02, 0.80) *	0.52 (0.01, 0.82) *
Channel 14	0.63 (-0.09, 0.79) *	0.62 (0.16, 0.86) *
Channel 15	0.20 (-0.51, 0.71)	0.18 (-0.44, 0.69)
Channel 16	0.61 (0.11, 0.86) *	0.59 (0.11, 0.85) *

Note. Significant values were shown with (*) $p < 0.05$.

across two versions (ICC(2,1) = 0.67; ICC(3,1) = 0.73, respectively). On the other hand, fair agreement and consistency were found for the reaction time during target letter detection, between two versions (ICC(2,1) = 0.46; ICC(3,1) = 0.59, respectively). ICC results for the alternate forms reliability indicated that behavioral measures were reliable over two forms of the same task (Table 32).

3.5.4 Alternate Forms Reliability of Hemodynamic Data

For the oxy-Hb measurements, fair agreement and consistency were found for Channel 3 (ICC(2,1) = 0.48; ICC(3,1) = 0.51), Channel 4 (ICC(2,1) = 0.52; ICC(3,1) = 0.54, Channel 5 (ICC(2,1) = 0.47; ICC(3,1) = 0.48), Channel 6 (ICC(2,1) = 0.52; ICC(3,1) = 0.53), Channel 7 (ICC(2,1) = 0.48; ICC(3,1) = 0.52), Channel 9 (ICC(2,1) = 0.50; ICC(3,1) = 0.49), Channel 14 (ICC(2,1) = 0.50; ICC(3,1) = 0.55). Good agreement and consistency were obtained for Channel 12 (ICC(2,1) = 0.65; ICC(3,1) = 0.71). Also, good agreement, but excellent consistency was found for Channel 10 (ICC(2,1) = 0.65; ICC(3,1) = 0.71). On the other hand, for Channel 8, 11, 13 and 16 no statistically significance reliability estimate was obtained and negative reliability estimates were calculated for Channel 1, 2, and 15. Results of the alternate form reliability study revealed reliability of oxy-Hb measurements across two versions of the same task (Table 33).

For the deoxy-Hb measures of the alternate form study, excellent agreement and consistency were found for Channel 7 (ICC(2,1) = 0.85; ICC(3,1) = 0.84). For Channel 8 (ICC(2,1) = 0.48; ICC(3,1) = 0.48), Channel 9 (ICC(2,1) = 0.59; ICC(3,1) = 0.57), Channel 11 (ICC(2,1) = 0.56; ICC(3,1) = 0.58) fair agreement and consistency were found. No statistical significance was obtained in Channel 3, 4, 5, 10, 12, 13, 15 and 16 and negative reliability estimates were observed in Channel 1, 2, 6, and 14, which signifies no reliability. Results demonstrated none reliability of deoxy-Hb measures over two versions of the same task, although some of the channels were found reliable, most of the channels were found none reliable (Table 34).

For the total-Hb measurements, fair agreement and consistency were found for Channel 3 (ICC(2,1) = 0.52; ICC(3,1) = 0.56), Channel 7 (ICC(2,1) = 0.51; ICC(3,1) = 0.52), Channel 10 (ICC(2,1) = 0.55; ICC(3,1) = 0.54). Good agreement and consistency was obtained for Channel 5 (ICC(2,1) = 0.63; ICC(3,1) = 0.62). For

Table 32

Intraclass Correlation Coefficients with 95% Confidence Interval for Behavioral Measurements of Alternate Forms Reliability Study

Behavioral Measurements	ICC(2,1) (95%-CI)	ICC(3,1) (95%-CI)
Target	0.55 (0.03, 0.82) *	0.55 (0.02, 0.84) *
Non-target	0.42 (-0.09, 0.77) *	0.55 (0.02, 0.81) *
Reaction Time Target	0.67 (0.18, 0.89) *	0.73 (0.32, 0.91) *
Reaction Time Non-Target	0.46 (-0.06, 0.80) *	0.59 (0.90,0.86) *

Note. Significant values were shown with (*) $p < 0.05$.

Table 33

Intraclass Correlation Coefficients with 95% Confidence Interval for oxy-Hb Measurements of Alternate Forms Reliability Study

Channels	ICC(2,1) (95%-CI)	ICC(3,1) (95%-CI)
Channel 1	0	0
Channel 2	0	0
Channel 3	0.48 (-0.02, 0.80) *	0.51 (-0.04, 0.82) *
Channel 4	0.52 (0.02, 0.82) *	0.54 (0.01, 0.83) *
Channel 5	0.47 (-0.07, 0.81) *	0.48 (-0.10, 0.82) *
Channel 6	0.52 (-0.01, 0.83) *	0.53 (-0.03, 0.84) *
Channel 7	0.48 (-0.04, 0.81) *	0.52 (-0.05, 0.83) *
Channel 8	0.26 (-0.19, 0.68)	0.32 (-0.28, 0.74)
Channel 9	0.50 (-0.04, 0.81) *	0.49 (-0.06, 0.81) *
Channel 10	0.73 (0.35, 0.91) *	0.75 (0.37, 0.92) *
Channel 11	0.45 (-0.11, 0.79)	0.44 (-0.12, 0.79)
Channel 12	0.65 (0.19, 0.88) *	0.71 (0.28, 0.90) *
Channel 13	0.39 (-0.13, 0.76)	0.41 (-0.16, 0.77)
Channel 14	0.50 (-0.01, 0.82) *	0.55 (0.00, 0.85) *
Channel 15	0	0
Channel 16	0.27 (-0.16, 0.67)	0.34 (-0.23, 0.74)

Note. Significant values were shown with (*) $p < 0.05$.

Table 34

Intraclass Correlation Coefficients with 95% Confidence Interval for deoxy-Hb Measurements of Alternate Forms Reliability Study

Channels	ICC(2,1) (95%-CI)	ICC(3,1) (95%-CI)
Channel 1	0	0
Channel 2	0	0
Channel 3	0.37 (-0.22, 0.75)	0.36 (-0.22, 0.75)
Channel 4	0.11 (-0.48, 0.61)	0.10 (-0.46, 0.60)
Channel 5	0.44 (-0.12, 0.81)	0.46 (-0.16, 0.82)
Channel 6	0	0
Channel 7	0.85 (0.55, 0.95) *	0.84 (0.53, 0.95) *
Channel 8	0.48 (-0.09, 0.82) *	0.48 (-0.10, 0.82) *
Channel 9	0.59 (0.06, 0.86) *	0.57 (0.06, 0.85) *
Channel 10	0.25 (-0.27, 0.68)	0.26 (-0.31, 0.70)
Channel 11	0.56 (0.07, 0.84) *	0.58 (0.06, 0.85) *
Channel 12	0.22 (-0.35, 0.67)	0.22 (-0.35, 0.67)
Channel 13	0.32 (-0.13, 0.71)	0.40 (-0.17, 0.77)
Channel 14	0	0
Channel 15	0.27 (-0.37, 0.73)	0.26 (-0.37, 0.73)
Channel 16	0.12 (-0.47, 0.62)	0.12 (-0.44, 0.61)

Note. Significant values were shown with (*) $p < 0.05$.

Channel 4, 8, 9, 11, 12, 13, 14, 15, and 16 no statistically significance reliability estimate was found. Negative reliability estimates were observed for Channel 1, 2, and 6. Results of the alternate forms reliability analysis showed that for most of the channels, total-Hb measures were not reliable across two versions of the same task (Table 35).

For the oxy-change measurements, fair agreement and consistency was found for Channel 3 (ICC(2,1) = 0.57; ICC(3,1) = 0.57). Good agreement and consistency were found for Channel 4 (ICC(2,1) = 0.65; ICC(3,1) = 0.70), Channel 6 (ICC(2,1) = 0.60; ICC(3,1) = 0.68), Channel 9 (ICC(2,1) = 0.70; ICC(3,1) = 0.71), and for Channel 13 (ICC(2,1) = 0.65; ICC(3,1) = 0.63). For Channel 14, fair agreement and good consistency were observed (ICC(2,1) = 0.57; ICC(3,1) = 0.65). Good agreement and excellent consistency were found for Channel 7 and Channel 10 (ICC(2,1) = 0.68; ICC(3,1) = 0.75; (ICC(2,1) = 0.73; ICC(3,1) = 0.79, respectively). Furthermore, excellent agreement and consistency were obtained for Channel 5 (ICC(2,1) = 0.84; ICC(3,1) = 0.84), Channel 11 (ICC(2,1) = 0.78; ICC(3,1) = 0.76) and for Channel 12 (ICC(2,1) = 0.78; ICC(3,1) = 0.86). However, for Channel 8, 15, and 16 no statistically significant reliability estimate was obtained. Also, for Channel 1 and 2, negative reliability estimates were found. Although there were not significant and negative estimates were found in the analysis, it can be said that overall findings of the alternate forms reliability indicated reliability of oxy-change measures between two versions of the same task (Table 36).

Table 35

Intraclass Correlation Coefficients with 95% Confidence Interval for total-Hb Measurements of Alternate Forms Reliability Study

Channels	ICC(2,1) (95%-CI)	ICC(3,1) (95%-CI)
Channel 1	0	0
Channel 2	0	0
Channel 3	0.52 (0.04, 0.82) *	0.56 (0.03, 0.84) *
Channel 4	0.29 (-0.29, 0.71)	0.28 (-0.29, 0.71)
Channel 5	0.63 (0.11, 0.88) *	0.62 (0.10, 0.87) *
Channel 6	0	0
Channel 7	0.51 (-0.02, 0.83) *	0.52 (-0.04, 0.83) *
Channel 8	0.27 (-0.22, 0.70)	0.31 (-0.29, 0.74)
Channel 9	0.44 (-0.13, 0.79)	0.43 (-0.13, 0.78)
Channel 10	0.55 (0.02, 0.84) *	0.54 (0.01, 0.83) *
Channel 11	0.29 (-0.26, 0.70)	0.29 (-0.28, 0.71)
Channel 12	0.38 (-0.16, 0.75)	0.39 (-0.18, 0.76)
Channel 13	0.23 (-0.22, 0.65)	0.27 (-0.31, 0.70)
Channel 14	0.25 (-0.32, 0.70)	0.26 (-0.35, 0.71)
Channel 15	0.41 (-0.11, 0.78)	0.48 (-0.14, 0.83)
Channel 16	0.18 (-0.27, 0.62)	0.21 (-0.36, 0.67)

Note. Significant values were shown with (*) $p < 0.05$.

Table 36

Intraclass Correlation Coefficients with 95% Confidence Interval for oxy-change Measurements of Alternate Forms Reliability Study

Channels	ICC(2,1) (95%-CI)	ICC(3,1) (95%-CI)
Channel 1	0	0
Channel 2	0	0
Channel 3	0.57 (0.08, 0.85) *	0.57 (0.06, 0.85) *
Channel 4	0.65 (0.20, 0.88) *	0.70 (0.26, 0.90) *
Channel 5	0.84 (0.55, 0.95) *	0.84 (0.54, 0.95) *
Channel 6	0.60 (0.08, 0.87) *	0.68 (0.20, 0.90) *
Channel 7	0.68 (0.17, 0.89) *	0.75 (0.36, 0.92) *
Channel 8	0.22 (-0.21, 0.65)	0.28 (-0.32, 0.72)
Channel 9	0.70 (0.29, 0.90) *	0.71 (0.28, 0.90) *
Channel 10	0.73 (0.27, 0.91) *	0.79 (0.44, 0.93) *
Channel 11	0.78 (0.41, 0.93)	0.76 (0.39, 0.92) *
Channel 12	0.78 (-0.16, 0.75)	0.86 (0.61, 0.96) *
Channel 13	0.65 (0.16, 0.88) *	0.63 (0.15, 0.87) *
Channel 14	0.57 (0.05, 0.85) *	0.65 (0.15, 0.88) *
Channel 15	0.27 (-0.26, 0.72)	0.30 (-0.33, 0.75)
Channel 16	0.31 (-0.14, 0.70)	0.38 (-0.19, 0.76)

Note. Significant values were shown with (*) $p < 0.05$.

CHAPTER 4

Discussion

While the human brain reacts to environmental stimuli, in the meantime it goes through some physiological changes. Those physiological changes occur as a result of our psychological experiences (Irani et al., 2007). As the technology progresses, various new brain imaging devices were developed. Those novel brain imaging devices are started to be widely used in order to determine those physiological changes more precisely and accurately to make conclusions about what is going on in the brain.

*f*NIR is one of the recently developed neuroimaging devices which allows researchers to measure hemodynamic changes during prefrontal cortex functioning. It is a safe, portable, affordable, non-invasive device and has higher temporal resolution compared to other brain imaging devices, such as *f*MRI and PET. Those characteristics of *f*NIR make it appropriate to be used in research and applied settings (Izzetoglu, 2008). It is developed from near-infrared spectroscopy (NIRS) based imaging systems which utilize optical methods to monitor oxy-Hb and deoxy-Hb concentrations in the tissue by using specific wavelengths (700-900 nm) in the near-infrared light spectrum. Within this spectrum, wavelengths of 700-900 nm are often referred to as the 'optical window' (Izzetoglu, 2008; Ferrari & Quaresima, 2012). With the help of light sources and photodetectors, the light that spreads from the light source passes through the scalp and frontal lobe tissues, and then held by the photodetectors. This light path that *f*NIR followed is known as "banana-shaped" path. The amount of absorbed and scattered light through the scalp are influenced by the hemodynamic changes that occur during the neural activity (Boas et al., 2001; Irani et al., 2007; Obrig et al., 1996; Obrig et al., 2000; Scholkmann et al., 2014; Strangman, Boas & Sutton, 2002). Those changes in the light intensity that are caused by neural activity are converted to oxy-Hb and deoxy-Hb concentrations by using modified Beer-Lambert Law (Cope, 1991; Cope & Delphy, 1997). Hence, the

concentration changes of oxy-Hb and deoxy-Hb occur, because of the hemodynamic activity changes. Those changes function based on neurovascular coupling (increase in oxy-Hb and decrease in deoxy-Hb concentrations). This mechanism is accepted as the indicator of hemodynamic activity in the brain and can be measured via *f*NIR device (Buxton et al., 2004; Obrig et al., 1996; Strangman et al., 2002; Plichta et al., 2006).

Within this context, every newly developed brain imaging device becomes a considerably valuable tool for researchers who are in a struggle for understanding and studying brain-mind-behavior relationship (Strangman, Boas & Sutton, 2002). On the other hand, those novel brain imaging devices are perceived as if they present the exact picture of the psychological experience. Many of the researchers accepted the information that this "high-technological systems" produce as accurate without questioning its validity and reliability. If the main purpose of brain imaging devices is to reflect the psychological experiences accurately and consistently, then this can possibly happen with initially determining the validity and reliability of measurements that the brain imaging devices produce. Otherwise, inferences which are drawn from the brain imaging studies cannot go beyond subjective interpretations of the brain imaging data. Therefore, the validity and reliability of the brain activity measurements emerge as an essential topic to study.

In psychometrics, validity refers to both the existing evidence for and the potential outcomes of test interpretation and usage. One type of validity is the construct validity. Construct validity is a central concept in all areas of psychology. In construct validity which is a type of validity now accepted by most to include all forms of validity, it is aimed to demonstrate the attributes or traits that the test was intended to measure and to determine the degree to which specific constructs account for performance on the test. In this thesis, the validity of the measurements was addressed based on the construct validity. To provide evidence to construct validity, it was expected to find similar results between behavioral and hemodynamic measures and the gender differences both on behavioral and hemodynamic measures was investigated to provide evidence to construct validity.

When the studies related to the validity of *f*NIR measures was examined, it is observed that most of the studies examine the validity of *f*NIR measures, by comparing the hemodynamic measurements obtained via *f*NIRS and *f*MRI (i.e., Cui

et al., 2011; Huppert et al., 2006; Kleinschmidt et al., 1996; Mehagnoul-Schipper et al., 2002; Murata et al., 2004; Strangman et al., 2002; Toronov et al., 2001; Sato et al., 2013). Moreover, in some of the studies, the validity of *f*NIR measures was examined by comparing the hemodynamic measures of healthy people and patients (i.e., Merzagora et al., 2008; Ruocco et al., 2007; Ayaz et al., 2014). Gender differences in behavioral performance and brain activity were one of the widely studied topics, although to investigate the construct validity was not the main reason of these studies, they can be accepted in the context of construct validity. The results of these studies based on gender differences showed contradictory behavioral and hemodynamic findings. While some researchers found differences in brain activity of males and females (Koch et al., 2007, Speck et al., 2010, Li et al., 2010; Gao, Zhang, Luo, Liu & Gong, 2016); some of them did not find gender differences in brain activations (Schmidt et al., 2009; Haut & Brach, 2006). On the other hand, researchers found similar behavioral performances among males and females (Li et al., 2010; Gao, Zhang, Luo, Liu & Gong, 2016; Schmidt et al., 2009; Haut & Brach, 2006). However, to investigate the construct validity was not the main reason for these studies.

Another psychometric characteristic which is known as the reliability refers to whether the device is free from measurement errors and measurements are consistent with repeated measurements under various conditions. Reliability of the measurements is assessed by internal consistency, alternate forms, and the test-retest method. In the investigation of the reliability of *f*NIRS measurements, generally, test-retest method was applied. To determine the test-retest reliability of measurements ICC is a generally used parameter (i.e., Watanabe, Matsuo, Kato, & Kato, 2003; Plichta et al., 2006; Plichta et al., 2012; Bhambhani, Maikala, Farag, & Rowland, 2015; Ruocco et al., 2007; Niu et al., 2013; Strangman, Goldstein, Rauch, & Stein, 2006; Wiggins, Anderson, Kitterick, & Hartley, 2016). Although consistent findings of behavioral measures across testing sessions were generally found (Schecklmann et al., 2008; Plichta et al., 2012; Bhambhani, Maikala, Farag, & Rowland, 2006); studies mostly found stable results at group level and at averaged measurements based on ROIs; but found highly variable results at single-subject level and single measurements (Schecklmann et al., 2008; Strangman et al., 2006; Wiggins, Anderson, Kitterick, & Hartley, 2016; Plichta et al; 2006). In this thesis, test-retest and alternate

forms reliability of measures were examined. Although test-retest method is a commonly used reliability method in *f*NIRS, alternate forms reliability was not investigated before. By considering the definition of reliability, the stability of both behavioral and hemodynamic measures was investigated among two testing sessions and two version of the same task. It was expected to find stable results between two testing session and two versions in both behavioral and hemodynamic measures.

As it was pointed out, with the increasing usage of *f*NIR device in the brain imaging studies, validity and reliability of the measurements obtained from *f*NIR device are started to be investigated. When the related literature is reviewed, it is observed that most of the studies relative to the validity of hemodynamic measurements obtained via *f*NIR device were investigated by comparing the results with *f*MRI and by comparing the hemodynamic measures of healthy people and patients. However, it is still inadequate. Because the validity of *f*NIRS measurements across Turkish sample was not determined in any of the studies, in spite of *f*NIR device is widely used in many studies conducted with Turkish samples. Therefore, it is inevitable to study the validity of *f*NIRS measurements across different populations. For the reliability of *f*NIRS measurements, researchers generally found stable results on behavioral measures across testing sessions, and also brain activity measures; but at the group level and averaged measurements across ROIs.

In this thesis, the construct validity of the behavioral and hemodynamic measurements was investigated based on gender differences by using linear mixed effects model approach. Because of the hierarchical structure of the data, in which repeated measures are nested within participants and its several advantages over conventional repeated measures analysis approaches, such as dealing with the homogeneity of variances, hierarchical sampling, and missing data, linear mixed model approach was used. Consistencies of both behavioral and hemodynamic measurements obtained via *f*NIRS were investigated by applying the reliability methods of test-retest (with the retest interval as 3 weeks) and alternate forms (different version of the same task). Measurements were obtained when participants were performing the n-back task, which is a widely utilized working memory task to examine the executive functions in the brain (Owen et al., 2005, Kane & Engle, 2002; Kane, Conway, Miura, & Colflesh, 2007). Letter version of n-back task, with four incremental cognitive loads (0-, 1-, 2-, and 3-back) was used as in the n-back studies

in the literature (Braver et al., 1997; Veltman, Rombouts, & Dolan, 2003; Izzetoğlu, Yurtsever, Bozkurt, & Bunce, 2003; Molteni et al., 2012; Sela, Izzetoğlu, Izzetoğlu, & Onaral; 2012; Ayaz, Izzetoglu, Bunce, Heiman-Patterson, & Onaral, 2012). N-back conditions were presented for three trials by conducting Latin Square block design to right-handed healthy participants. While participants were performing the n-back task, changes in their prefrontal cortex were measured with the 16-channeled *f*NIR device. Behavioral results were examined based upon the accuracy of the responses and reaction time. Responses of the participants were evaluated as accurate if they responded to a target letter as target letter (by pressing left click) and responded to a non-target letter as non-target letter (by pressing right click) based on the instruction. Reaction times during the target and non-target letter detection were also calculated from the beginning until the end of that n-back condition. Hemodynamic measures were investigated upon oxy-Hb, deoxy-Hb, total-Hb and oxygenation change measures.

Findings of the behavioral results revealed that as the n-back conditions increased from 0- to 3-back, the number of detected target and non-target letters decreased; but response times on target and non-target detection increased, compared with the baseline condition (0-back condition in the first trial). Results of the accuracy of the responses and reaction time of the participants revealed that cognitive workload impacted both accuracy and reaction time during performance of participants, which is consistent with the findings of other n-back studies in the literature (Veltman, Rombouts, & Dolan, 2003; Ayaz et al., 2012; Izzetoğlu, 2008; Schmidt et al., 2009; Molteni et al., 2012, Sela, Izzetoğlu, Izzetoğlu, & Onaral; 2012; Li, Luo & Gong, 2010). It can be implied that as the cognitive task demands increase, individuals started to respond less accurately and more slowly. This signifies the decline of working memory performance as a result of high cognitive demand. On the other hand, it was found that detected non-target letters increased in 3-back as the trials progressed; and reaction time during target and non-target letter detection decreased in 3-back in the third trial, compared with the baseline condition. One possible reason for this finding is may be due to the high cognitive demand of the task. Individuals may give up to the task because of the difficulty of 3-back condition and started to give immediate responses, hence some of their responses lead to accurate responses in non-target letters. This behavioral finding indicated that, as it

was hypothesized, before considering gender-specific differences on behavioral measurements, participants started to respond less accurately and more slowly as the working memory load increased.

Results of the hemodynamic measures indicated that for oxy-Hb measures, an initial decrease of oxy-Hb measures in 3-back condition compared with the baseline condition which is 0-back condition in the first trial and was found only in Channel 2, which corresponds to left ventrolateral prefrontal cortex (VLPFC) (Ayaz et al., 2012). However, oxy-Hb measures in 3-back condition were started to increase in the second trial, compared with the baseline. Significant hemodynamic activity only in Channel 2 was also found by Ayaz and his colleagues (2012). They demonstrated increased activity in 3-back condition compared with the 0-back condition and 1-back condition, only in Channel 2, but only the reported the oxygenation change. For the deoxy-Hb measures, an initial increase of deoxy-Hb in 3-back condition in Channel 1 (left DLPFC), Channel 2 (left VLPFC), and Channel 16 (right VLPFC); in 2-and 3-back in Channel 15 (right VLPFC) were found compared with the 0-back condition in the first trial. Increased deoxy-Hb measures were also observed in the second trial in Channel 8 (bilateral ventromedial PFC) and Channel 11 (right dorsomedial PFC); in the second trial and in the third trial in Channel 4 (left VLPFC), Channel 6 (left middle frontal gyrus), Channel 9 (right dorsomedial PFC), Channel 12 (right lateral frontopolar cortex), Channel 14, Channel 15 and Channel 16 (right VLPFC) as compared with the 0-back condition in the first trial. However, deoxy-Hb measures started to decrease in the second and in the third trial for 2-back and 3-back conditions when compared with the 0-back condition in the first trial, except in Channel 9 and Channel 11 (right dorsomedial PFC). In Channel 9 and Channel 11, no significant decreases in deoxy-Hb measures were observed. Apart from all channels in deoxy-Hb measures of Channel 2, decreases in deoxy-Hb measures were also found in 1-back condition in the second and in the third trial. For total-Hb measures, no significant activation was found in any of the channels. This result seems not to support the related literature; because total-Hb measures were found statistically significant in most of the studies. Total-Hb was calculated from the sum of oxy-Hb and deoxy-Hb measures, and indicates the cerebral blood volume (Mehagnoul-Schipper et al., 2002). Therefore, changes in total-Hb can be used as a measure of changes in blood volume (Delpy et at., 1998; Berner, 2015; Schecklmann

et al., 2008). Due to the low spatial resolution of *f*NIRS (1.25 cm penetration depth), it may restrict the signals to a small cerebral volume under the channels; thus CBV of the participants may not be detected and no significant activity was found in total-Hb measures as Gratton and his colleagues suggested (1994). For the oxygenation change measures, decreased oxygenation change measures was observed in second and third trial in Channel 1 and in Channel 4 (left DLPFC), in second trial Channel 8 (bilateral ventromedial PFC) and in Channel 12 (right lateral frontopolar cortex), Channel 14, and Channel 15 (right VLPFC). Additionally, in Channel 2 (left VLPFC) decreased oxygenation change in 3-back condition, in the second and in the third trial; and increased oxygenation change in 2-back in the second trial, 3-back in the second trial, and 3-back in the third trial was observed, compared with the baseline condition. Overall results of the hemodynamic measures were found consistent with the findings of the *f*MRI study of Braver and his colleagues (1997) in which a verbal n-back with increasing loads from 0-, 1-, 2-, and 3-back was used. They found cognitive workload dependent activation in bilateral DLPFC and left VLPFC. In another *f*MRI, which is conducted by Veltman, Rombouts, and Dolan (2003), again a verbal n-back task with loads of 0-, 1-, 2-, and 3-back was presented and bilateral DLPFC, left VLPFC, and left parietal cortex activation depended on cognitive workload was observed. As it was hypothesized, this hemodynamic finding showed that, before investigating gender-specific differences on hemodynamic measurements, participants' oxy-hemoglobin and oxygenation measurements increased and deoxy-hemoglobin decreased as the working memory load and trials increased.

As it was initially mentioned, increased neural activity is accepted as an increase in oxy-Hb and decrease in deoxy-Hb measurements. However, in this thesis oxy-Hb measurements decreased and then increased in 3-back condition and deoxy-Hb measurements increased and then decreased as the trials progressed. This finding may be explained by the "initial dip" phenomenon. This phenomenon states an early deoxy-Hb increase followed by a later and more noticeable decrease in deoxy-Hb as a proof for early desaturation of hemoglobin and at the initiation of functional activation. This situation happens because of the increased oxygen metabolism prior to the increase of cerebral blood flow, results in a dynamic uncoupling of blood flow and metabolism (Buxton, 2001; Shepro, 2006).

For the construct validity of measures, similar results between behavioral and hemodynamic measures in terms of gender difference was found. Males and females did not differ on behavioral measures, in accuracy and reaction time during non-target detection; and they did not differ in oxy-Hb, deoxy-Hb, and oxygenation change measurements, except in Channel 14 (right VLPFC) in deoxy-Hb measures. As it was hypothesized, similar finding of behavioral and hemodynamic measures with respect to gender differences was accepted as a proof for the construct validity of measures obtained via *f*NIRS. Those findings are almost identical to the results of the *f*MRI study of Schmidt and her colleagues (2009). They found that males and females performed similarly on behavioral measures and have similar hemodynamic activity measures. For behavioral measures, they found that both males and females had decreased accuracy and increased reaction time as the cognitive work load increased. For hemodynamic activity measures, they found bilateral activation in the superior frontal gyrus (BA 6), middle frontal gyrus (BA 9 and 10), inferior frontal gyrus (BA 47) and the inferior parietal lobule (BA 40) in both genders. On the other hand, results of the construct validity measures of this thesis demonstrated gender differences in reaction time during target letter detection in 2-back and 3-back conditions in the second trial. It was observed that females' reaction time during target letter detection was faster than males' reaction time during target letter detection in 0-back condition in the first trial. This result was inconsistent with the findings of Speck and his colleagues (2009) who found females have slightly slower reaction time compared to males. Although same working memory task with Speck and his colleagues was used in this thesis, there are some differences. They presented letter and number version of 1-back and 2-back conditions to 9 male and 8 females twice; however, in this study, each four n-back conditions were presented three times and the sample size was larger compared to their study. Therefore, it may not be entirely comparable with the result of Speck and his colleagues. One explanation to find gender difference in reaction time during target letter detection is may be because gender in itself did not result in differences in reaction time during target letter detection; the significant finding of interaction may come from the significant results of 2-back and 3-back conditions. Nevertheless, similar to the behavioral measurements, males and females did not differ on hemodynamic activity measures; specifically on oxy-Hb, oxygenation change, and deoxy-Hb measures. Except in

Channel 14, that corresponds to right VLPFC. Females showed increased deoxy-Hb measures in the 1-back condition; but decreased deoxy-Hb measures in the 1-back condition in the second and in the third trial, compared with the 0-back condition in the first trial of males. The result of the decreased of deoxy-Hb measures of females in the right VLPFC was not consistent with the findings of Speck and his colleagues (2009), in which their *f*MRI data results showed predominant left hemisphere activation of females. Also, Li and her colleagues (2010) found left-lateralized activation of oxy-Hb and total-Hb measures. One possible reason to find right hemisphere activation is that females themselves did not differ, but with the significant decrease in 1-back condition, they show variations. Therefore this significant finding of females may come from the significant result of 1-back.

For the test-retest reliability of behavioral measures, good to excellent agreement and consistency between two testing sessions was found. Although poor agreement and fair consistencies were found in accuracy and reaction time of non-target letters, results were found statistically significant. Therefore, test-retest findings revealed stability of behavioral measures across two testing sessions. This result is consistent with the behavioral results of Plichta and his colleagues (2012), in which their behavioral measurements (face matching, monetary reward anticipation, and n-back task) were stable across testing sessions for most of the participants.

The test-retest reliability of *f*NIR measures was investigated based on single measurement and single-subject level. For oxy-Hb, fair to good agreement and consistency for Channel 2, 3, 4, 6, 7, 8, 9, 10, 12, 14 and 16 were found. For the deoxy-Hb measures, fair agreement and consistency were only observed in Channel 7. For the total-Hb measures, fair to excellent agreement and good consistency for Channel 2, 3, 15 were detected. For the oxygenation change measures, fair to excellent agreement and consistency for Channel 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, and 16 were found.

For the alternate forms reliability of behavioral measures, fair to good agreement and consistency between two versions of the test was found. For the oxy-Hb measures, fair to excellent agreement and consistencies were observed for Channel 3, 4, 5, 6, 7, 9, 10, 12, and 14. For the deoxy-Hb measures, fair to excellent agreement and consistency were observed only for Channels 7, 8, 9 and 11. For the total-Hb measures, fair to good agreement and consistency were obtained for

Channel 3, 5, 7, and 10. Lastly, for oxygenation change measures, fair to excellent agreement and consistency were observed in Channels 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, and 14. From overall test-retest and alternate form findings, it can be concluded that oxy-Hb and oxygenation measures were stable; but deoxy-Hb and total-Hb measures were not so stable across two versions. However, those test-retest findings of *f*NIR measures seem not in agreement with the literature; because studies generally found stable results at a group level and averaged measurements across multiple channels covering a cortical ROI; but found highly variable results at the single-subject level and single measurement. Therefore, researchers recommended to use group level and averaged measurements to the specified ROI (Schecklmann et al., 2008; Strangman et al., 2006; Wiggins et al., 2016; Plichta et al; 2006). In spite of that, researchers should keep in their mind that, *f*NIR device measures hemodynamic activities from the head surface via sensor pad. Those pads were positioned in reference to standard 10–20 coordinates. Therefore, the location of channels related to underlying brain regions could not be independently determined. Also, positioning channels over a specific cortical region hinge upon external anatomical markers. Thus, these factors make precise localization of *f*NIRS signals difficult (Fishburn et al., 2014; Okamoto et al., 2004; Mehagnoul-Schipper et al., 2002).

Throughout this thesis, several limitations were encountered. One of the limitations is the loss of participants. Because *f*NIR is susceptible of different contours, skin colors and skin/skull thickness, some of the participants could not adapt to *f*NIR device; therefore, could not take part in the study. In addition, because the raw light is sensitive to hair, which is not able to travel through hair, hair on some of the regions of the participants' forehead results in noise in the data. This circumstance sometimes causes loss of the channels, after noise removal algorithms. In addition to this, *f*NIRS is highly susceptible to motion artefacts, although it was required from the participant to preserve their positions throughout the study, any movements influence the signal quality and also the position of the sensor pad. Therefore, sometimes this leads removal of the whole data. Additionally, due to low spatial resolution and low penetration depth into the brain regions (approximately 1.25 cm), it may restrict the signals to a small cerebral volume underneath the channels. Hence, poor and/or no activation may be obtained. Apart from *f*NIR device, some limitations may occur due to the experimental design. Even though *f*NIRS has

high temporal resolution (500 msec per scan), and n-back types were presented three times to participants; it may be better to extent time of conditions by increasing the number of presented letters in order to increase the quantity of measurements.

For future studies, it is suggested to administer different versions of n-back tasks, such as number version and/or shape/figure version. Also, apart from the n-back task, other tasks which are related to executive functions of the brain can be applied, such as Wisconsin Card Sorting test, and/ or Trail Making Test, and/or Stroop Test, in order to see whether using different working memory or different versions of the same tasks lead differences in behavioral and hemodynamic measures. To get similar results by using different tests will be another evidence for the validity of hemodynamic measures. Moreover, hemispheric differences can be investigated based on right or left handedness or by presenting stimuli from left or right-hand side. By considering lateralization, interpretations of the significant channels may become more straightforward. Furthermore, for the validity of *f*NIR measures, construct validity study can be conducted by comparing measurements obtained from a healthy and patient sample selected from the Turkish population. Additionally, for the test-retest reliability of *f*NIR measures, the sample size may increase, in order to generalize the consistency of measures across a wider sample.

All in all, as Logothetis (2008) emphasized there is an increasing trend in the literature that reinforces the integrated views of the brain functioning. As a consequence, it was thought that there will be more evidence for the validity of the measures obtained from devices if brain imaging is approached with the collaboration of knowledge about the psychological and cognitive states, experimental and statistical methods in psychology; technological facilities in software and engineering; and knowledge about the underlying neurobiological systems and mechanisms in neuroscience. In this thesis, this suggested integrated approach was tried to be applied; but it sheds light on only a small part of this immense area. Hence, more studies with the integrated viewpoint should be conducted.

References

Abdelnour, F., Schmidt, B., & Huppert, T. J. (2009). Topographic localization of brain activation in diffuse optical imaging using spherical wavelets. *Physics in Medicine and Biology*, *54*, 6383–6413.

Akaike, H. (1977). On entropy maximization principle. In Krishnaiah PR (Ed.), *Applications of statistics*. Amsterdam: North-Holland (pp. 27–41).

Ames, A. (2000). CNS energy metabolism as related to function. *Brain Research Reviews*, *34*, 42–68.

Anastasi, A. (1982). Psychological testing (5th ed.). New York: Macmillan.

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science*, *7*, 25–31.

Ayaz, H. (2010). Functional near-infrared spectroscopy based brain computer interface. *PhD Thesis*, Drexel University, Philadelphia, PA.

Ayaz, H., Izzetoglu, M., Bunce, S. C., Heiman-Patterson, T., Onaral, B. (2007) Detecting cognitive activity related hemodynamic signal for brain computer interface using functional near infrared spectroscopy. *3rd IEEE Neuroengineering Conference, Hawaii,* 342-345.

Ayaz, H., Izzetoglu, M., Platek, S. M., Bunce, S., Izzetoglu, K., Pourrezaei, K., & Onaral, B. (2006). Registering *f*NIR data to brain surface image using MRI templates. *28th The IEEE Engineering in Medicine and Biology Society Annual International Conference, New Yor*k.

Ayaz, H., Izzetoglu, M., Shewokis, P. A., & Onaral, B. (2010). Sliding-window motion artifact rejection for functional near-infrared spectroscopy. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*.

Ayaz, H., Shewokis, P. A, Curtin, A., Izzetoglu, M., Izzetoglu, K., & Onaral, B. (2011). Using MazeSuite and functional near infrared spectroscopy to study learning in spatial navigation. *Journal of Visualized Experiments : JoVE*, 1–12.

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2012b). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage*, *59*, 36–47.

Ayaz, H., Shewokis, P. A., Scull, L., Libon, D. J., Feldman, S., Eppig, J., Onaral, B., Heiman-Patterson, T. (2014). Assessment of Prefrontal Cortex Activity in Amyotrophic Lateral Sclerosis Patients with Functional Near Infrared Spectroscopy. *Journal of Neuroscience and Neuroengineering*, *3*(1), 41-51.

Bartocci, M., Winberg, J., Ruggiero, C., Bergqvist, L. L., Serra, G., & Lagercrantz, H. (2000). Activation of olfactory cortex in newborn infants after odor stimulation: A functional near-infrared spectroscopy study. *Pediatric Research*, *48*, 18–23.

Bates, D., Maechler M., & Bolker, B. (2013). lme4: Linear-mixed Effects Models Using S4 Classes. Retrieved from: http://CRAN.R-project.org/package=lme4 (R package version 0.999999-2).

Beachey, W. (2012). Oxygen equilibrium and transport. *Respiratory care anatomy and physiology: Foundations for clinical practice* (3rd ed., pp. 143-146). Elsevier Inc.

Bélanger, M., Allaman, I., & Magistretti, P. J. (2011). Brain energy metabolism: Focus on astrocyte-neuron metabolic cooperation. *Cell Metabolism*, *14*, 724–738.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289-300.

Berner, A. L. (2015). Self-regulatory control of eating: An *f*NIR study of bulimia nervosa, *PhD Thesis*, Drexel University, Philadelphia, PA.

Bhambhani, Y., Maikala, R., Farag, M., & Rowland, G. (2015). Reliability of near-infrared spectroscopy measures of cerebral oxygenation and blood volume during handgrip exercise in nondisabled and traumatic brain-injured subjects. *Journal of Rehabilitation Research and Development*, *43*, 845–856.

Boas, D. A., Franceschini, M. A., Dunn, A. K., & Strangman, G. (2002).Non-Invasive imaging of cerebral activation with diffuse optical tomography. In R.

D. Frostig (Ed.), *In-vivo optical imaging of brain function* (pp. 193–221). CRC Press.

Boas, D. A., Gaudette, T., Strangman, G., Cheng, X., Marota, J. J. A., & Mandeville, J. B. (2001). The accuracy of near infrared spectroscopy and imaging during focal changes in cerebral hemodynamics. *NeuroImage*, *13*, 76–90.

Boas, D. A., Strangman, G., Culver, J. P., Hoge, R. D., Jasdzewski, G., Poldrack, R. A., … Mandeville, J. B. (2003). Can the cerebral metabolic rate of oxygen be estimated with near-infrared spectroscopy? *Physics in Medicine and Biology*, *48*, 2405–2418.

Bonelli, R. M., & Cummings, J. L. (2007). Frontal-subcortical circuitry and behavior. *Dialogues in Clinical Neuroscience*, *9*, 141–151.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory, *Neuroimage*, *5*, 49–62.

Brazy, J. E., Lewis, D. V., Mitnick, M. H., & Jöbsis vander Vliet, F.F. (1985). Noninvasive monitoring of cerebral oxygenation in preterm infants: preliminary observations. *Pediatrics*, *75*, 217–225.

Broca, P. (1861). Remarques sur le siège de la faculté du langage articule suivies d'une observation d'aphemie. *Bulletins de la Société Anatomique de Paris*, *6*, 330357.

Brodmann, K. (1909). Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. In L. J. Garey (Trans. &Ed.), Brodmann's Localization in the Cerebral Cortex. *Description of individual brain maps* (pp. 131-170). United States of America: Springer.

Bryk, A. S. & Raudenbush, S. W. (2002). Hierarchical Linear Models: Applications and Data Analysis Methods. *Advanced Quantitative Techniques in the Social Sciences* (2nd ed.). Sage Publications, Ltd.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*(l), 147–158.

Bunge, S.A. & Kahn, I. (2009). Cognition: An Overview of Neuroimaging Techniques. In L. R. Squire (Ed.), *Encyclopedia of neuroscience* (pp. 1063–1067). Oxford: Academic Press. Elsevier.

Buxton, R. B. (2001). The elusive initial dip. *NeuroImage*, 13, 953–958.

Buxton, R. B., Uludag, K., Dubowitz, D., Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *NeuroImage*, *23*, 220–233.

Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magnetic Resonance in Medicine*, *39*, 855–864.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Canning, C., & Scheutz, M. (2013). Functional near-infrared spectroscopy in human-robot interaction. *Journal of Human-Robot Interaction*, *2*, 62–84.

Carlson, S., Martinkauppi, S., Rämä, P., Salli, E., Korvenoja, A., & Aronen, H. J. (1998). Distribution of cortical activation during visuospatial n-back tasks as revealed by functional magnetic resonance imaging. *Cerebral Cortex*, *8*, 743–752.

Carmines, E. G., & Zeller, R. A. (1979). Reliability and validity assessment. *Sage university papers series. Quantitative applications in the social sciences.* 1-71.

Chan, D. (1998a). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods*, *1*, 421-483.

Chance, B., Anday, E., Nioka, S., Zhou, S., Hong, L., Worden, K., … Thomas, R. (1998). A novel method for fast imaging of brain function, noninvasively, with light. *Optics Express*, *2*, 10.

Chance, B., Zhuang, Z., UnAh, C., Alter, C., & Lipton, L. (1993). Cognition-activated low frequency modulation of light absorption in human brain. *Proceedings of the National. Academy of Sciences of the United States of America*, *90*, 3770–3774.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, *386*, 604–608.

Coolidge, F. L., & Wynn, T. (2005). Working memory, its executive functions, and the emergence of modern thinking. *Cambridge Archaeological Journal*, *15*, 5–26.

Cope, M. (1991). The development of a near-infrared spectroscopy system and its application for noninvasive monitoring of cerebral blood and tissue oxygenation in the newborn infant. *PhD Thesis*, University of London.

Cope, M., & Delpy, D. T. (1988). System for long term measurement of cerebral blood and tissue oxygenation on newborn infants by near infrared transillumination. *Medical & Biological Engineering Computing*, *26*, 289–294.

Crocker, L., and Algina, J. (1986). Introduction to Classical and Modern Test Theory. New York: CBS College Publishing.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological test. *Psychological Bulletin*, *52*, 281–302. https://doi.org/10.1037/h0040957

Csibra, G., Henty, J., Volein, Á., Elwell, C., Tucker, L., Meek, J., & Johnson, M. H. (2004). Near infrared spectroscopy reveals neural activation during face perception in infants and adults. *Journal of Pediatric Neurology*, *2*, 85–89.

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., & Reiss, A. L. (2010). A quantitative comparison of NIRS and *f*MRI across multiple cognitive tasks. *NeuroImage*, *54*, 2808–2821.

D'Esposito, M., Detre, J. A, Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature*, *378*, 279–81.

Dade, L. A., Zatorre, R. J., Evans, A. C., & Jones-Gotman, M. (2001). Working memory in another dimension: functional imaging of human olfactory working memory. *NeuroImage*, *14*, 650–660.

Delpy, D. T., & Cope, M. (1997). Quantification in tissue near-infrared spectroscopy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *352*(1354), 649–659.

Delpy, D. T., Cope, M., Cady, E. B., Wyatt, J. S., Hamilton. P. A., Hope, P. L., … Reynolds, E. O. (1987). Cerebral monitoring in newborn infants by magnetic resonance and near-infrared spectroscopy. *Scandinavian Journal of Clinical Laboratory Investigation*, *188*, 9–17.

Druzgal, T. J., & D'Esposito, M. (2001). Activity in fusiform face area modulated as a function of working memory load. *Cognitive Brain Research*, *10*, 355–364.

Erkuş, A. (2003). Geçerlik. *Psikometri üzerine yazılar: Ölçme ve psikometrinin tarihsel kökenleri, güvenirlik, geçerlik, madde analizi, tutumlar, bileşenleri ve ölçülmesi* (pp. 74). Ankara: Türk Psikologlar Derneği Yayınları.

Erkuş, A. (2003). Güvenirlik. *Psikometri üzerine yazılar: Ölçme ve psikometrinin tarihsel kökenleri, güvenirlik, geçerlik, madde analizi, tutumlar, bileşenleri ve ölçülmesi* (pp. 36). Ankara: Türk Psikologlar Derneği Yayınları.

Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (*f*NIRS): Development and fields of application. *Neuroimage*, *63*, 921–935.

Ferrari, M., Giannini, I., Carpi, A., & Fasella, P. (1982). Near-infrared spectroscopy in non-invasive monitoring of cerebral function. *World Congress on Medical Physics and Biomedical Engineering*, *Hamburg*.

Ferrari, M., Giannini, I., Carpi, A., Fasella, P., Fieschi, C., & Zanette, E. (1980). Non-invasive infrared monitoring of tissue oxygenation and circulatory parameters. *XII World Congress of Angiology*, *Athens*.

Ferrari, M., Giannini, I., Sideri, G., & Zanette, E. (1985). Continuous non-invasive monitoring of human brain by near infrared spectroscopy. *Advances in Experimental Medicine and. Biology*, *191*, 873–882.

Ferrari, M., Mottola, L., & Quaresima, V. (2004). Principles, techniques, and limitations of near-infrared spectroscopy. *Canadian Journal of Applied Physiology*, *29*, 463–487.

Field, A. (2009). Why is my evil lecturer forcing me to learn statistics?. In M. Carmichael (Ed.), *Discovering Statistics Using IBM SPSS Statistics* (3rd ed.,pp. 10–13). SAGE Publications Ltd.

Field, A., Miles, J., & Field, Z. (2012). Multilevel linear models. *Discovering Statistics Using R* (pp. 997–1050). SAGE Publications Ltd.

Fishburn, F. A, Norr, M. E., Medvedev, A. V, & Vaidya, C. J. (2014). Sensitivity of *f*NIRS to cognitive state and load. *Frontiers in Human Neuroscience*, *8*, 76.

Fox , P. T. , Raichle , M. E. , Mintun ,M. A. , & Dence , C. (1988). Nonoxidative glucose consumption during focal physiologic neural activity. *Science*, *241* (4864), 462–464.

Franceschini M. A., Fantini S., Thompson J. H., Culver J. P., Boas D. A. (2003). Hemodynamic evoked response of the sensorimotor cortex measured noninvasively with near-infrared optical imaging. Psychophysiology 40, 548–560.

Franzen, M. D. (2002) Reliability and validity in neuropsychological assessment, In A. E. Puente & C. R. Reynolds (Eds.), *Critical issues in neuropsychology* (2nd ed.). Springer.

Fuchs, A. F & Phillips, J. O. (1989). Association cortex. In H. D. Patton, A. F. Fuchs, B. Hille, A. M. Scher, R. Steiner (Eds.), *Textbook of physiology* (21st ed., pp. 663–692). United States of America: W. B. Saunders Company.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, *61*, 331–349.

Furr, R.M., & Bacharach, V.R. (2014). Psychometrics: An introduction (2nd ed.) Thousand Oaks, CA: Sage.

Galecki, A., & Burzykowski, T. (2013). Linear mixed-effects model. In G. Casella, S.E. Fienberg & I. Olkin (Eds). *Linear Mixed-Effects Models Using R: A Step by Step Approach* (pp. 245-326). Springer.

Gao C., Zhang L., Luo D., Liu D., Gong H. (2016) PFC activity pattern during verbal WM task in healthy male and female subjects: A NIRS study. In Luo

Q., Li L., Harrison D., Shi H., Bruley D. (Eds). *Oxygen Transport to Tissue XXXVIII. Advances in Experimental Medicine and Biology*, 923. Springer.

Giannini, I., Ferrari, M., Carpi, A., & Fasella, P. (1982). Rat brain monitoring by near-infrared spectroscopy: an assessment of possible clinical significance. *Physiological Chemistry and Physics*, *14*, 295–305.

Gratton, G., Corballis, P. M., Cho, E., Fabiani, M., & Hood, D. C. (1995). Shades of gray matter: noninvasive optical images of human brain responses during visual stimulation. *Psychophysiology*, *32*, 505–9.

Gulliksen, H. (1950). Intrinsic validity. *American Psychologist*, *5*(10), 511–517.

Hardcastle, V., & Stewart, C. M. (2002). What Do Brain Data Really how?. *Philosophy of Science*, 69, 72-82.

Harlow J. M. (1848). Passage of an iron rod through the head. *Boston Medical and Surgical Journal, 39*, 389–393

Harlow, J. M. (1868). Recovery after Severe Injury to the Head. *Bulletin of the Massachusetts Medical Society*. *4*, 274–281.

Haut, K. M., & Barch, D. M. (2006). Sex influences on material-sensitive functional lateralization in working and episodic memory: men and women are not all that different. *Neuroimage*, *32*, 411–422.

Hautzel, H., Mottaghy, F. M., Schmidt, D., Zemb, M., Shah, N. J, Muller-Gartner, H. W, & Krause, B. J. (2002). Topographic segregation and convergence of verbal, object, shape and spatial working memory in humans. *Neuroscience Letters*, *323*, 156–160.

Healy, M. (2001). Multilevel data and their analysis. In A. H. Leyland & H. Goldstein, H (Eds.), *Multilevel modeling of health statistics*. (pp 1-12): Wiley; West Sussex.

Heinzel, S., Haeussinger, F. B., Hahn, T., Ehlis, A. C., Plichta, M. M., & Fallgatter, A. J. (2013). Variability of (functional) hemodynamics as measured with simultaneous *f*NIRS and *f*MRI during intertemporal choice. *NeuroImage*, *71*, 125–134.

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., & Schultz, T. (2014). Mental workload during n-back task-quantified in the prefrontal cortexusing *f*NIRS. *Frontiers in Human Neuroscience*, *7*, 935.

Hernandez-Meza, G., Slason, L., Ayaz, H., Craven, P., Oden, K., & Izzetoğlu., K. (2015). Investigation of functional near infrared spectroscopy in evaluation of pilot expertise acquisition. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Foundations of augmented cognition* (pp. 232–243): Springer International Publishing.

Hiroyasu, T., Nakamura, Y. Y., & Yokouchi, H. (2013). Method for removing motion artifacts from *f*NIRS data using ICA and an acceleration sensor. *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 6800–6803.

Hock, C., Villringer, K., Heekeren, H., Hofmann, M., Wenzel, R., Villringer, A., Müller-Spahn, F. (1997). A role for near infrared spectroscopy in psychiatry? *Advances in Experimental Medicine and Biology*, *413*, 105–123.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: foundations and illustrative examples. *Behavior Research Methods*, *39*(1), 101–117.

Hoge, R. D., Franceschini, M. A., Covolan, R. J. M., Huppert, T., Mandeville, J. B., & Boas, D. A. (2005). Simultaneous recording of task-induced changes in blood oxygenation, volume, and flow using diffuse optical imaging and arterial spin-labeling MRI. *NeuroImage*, *25*, 701-707.

Honey, G. D., Bullmore, E. T., & Sharma, T. (2000). Prolonged reaction time to a verbal working memory task predicts increased power of posterior parietal cortical activation. *NeuroImage*, *12*, 495–503.

Hoshi, Y. (2003). Functional near-infrared optical imaging: Utility and limitations in human brain mapping. *Psychophysiology*, *40*, 511–520.

Hoshi, Y., & Tamura, M. (1993a). Detection of dynamic changes in cerebral oxygenation coupled to neuronal function mental work in man. *Neuroscience Letters*, *150*, 5–8.

Hoshi, Y., & Tamura, M. (1993b). Dynamic multichannel near-infrared optical imaging of human brain activity. *Journal of Applied Physiology*, *75*, 1842–1846.

Hox, J. (2010). Multilevel analysis techniques and applications. In G. A. Marcoulides (Ed.), *Introduction to Multilevel Analysis* (pp. 1-32). New York: NY.

Hox, J. Multilevel analysis: Techniques and applications. Lawrence Erlbaum; Mahwah, NJ: 2002.

Huppert, T. J., Diamond, S. G., & Boas, D. A. (2008). Direct estimation of evoked hemoglobin changes by multimodality fusion imaging. *Journal Biomedical Optics*, *13*, 054031.

Huppert, T. J., Hoge, R. D., Dale, A. M., Franceschini, M. A., & Boas, D. A. (2006a). Quantitative spatial comparison of diffuse optical imaging with blood oxygen level-dependent and arterial spin labeling-based functional magnetic resonance imaging. *Journal of Biomedical Optics*, *11*, 064018.

Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2006b). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage*, *29*, 368–82.

Irani, F., Platek, S. M., Bunce, S., Ruocco, A. C., & Chute, D. (2007). Functional near infrared spectroscopy (*f*NIRS): An emerging neuroimaging technology with important applications for the study of brain disorders. *The Clinical Neuropsychologist*, *21*, 9–37.

Izzetoglu, K. (2008). Neural Correlates of Cognitive Workload and Anesthetic Depth: *f*NIR Spectroscopy Investigation in Humans. *Phd Thesis*, Drexel University.

Izzetoglu, K., Ayaz, H., Merzagora, A., Izzetoglu, M., Shewokis, P. A., Bunce, S. C., … Onaral, B. (2011). The evolution of field deployable *f*NIR spectroscopy from bench to clinical settings. *Journal Innovative Optical Health Sciences*, *4*, 1–12.

Izzetoglu, K., Bunce, S. C., Pourrezaei, K., & Chance, B. (2004). Functional optical brain imaging using NIR during cognitive tasks. *International Journal of Human– Computer Interaction*, *17*, 211–227.

Izzetoglu, K., Bunce, S., Izzetoglu, M., Onaral, B., & Pourrezaei, K. (2003). fNIR spectroscopy as a measure of cognitive task load. *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 4*, 3431–3434.

Izzetoglu, M., Bunce, S. C., Izzetoglu, K., Onaral, B.,& Pourrezaei, K. (2007). Functional brain imaging using near-infrared technology. *IEEE Engineering in Medicine and Biology Magazine, 26*, 38–46.

Izzetoglu, M., Chitrapu, P., Bunce, S., & Onaral, B. (2010). Motion artifact cancellation in NIR spectroscopy using discrete Kalman filtering. *Biomedical Engineering Online, 9*, 16. https://doi.org/10.1186/1475-925X-9-16

Izzetoglu, M., Izzetoglu, K., Bunce, S., Ayaz, H., Devaraj, A., Onaral, B., & Pourrezaei, K. (2005). Functional near-infrared neuroimaging. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 13*(2), 153–159.

Jaeggi, S. M., Seewer, R., Nirkko, A. C., Eckstein, D., Schroth, G., Groner, R., & Gutbrod, K. (2003). Does excessive memory load attenuate activation in the prefrontal cortex? Load-dependent processing in single and dual tasks: Functional magnetic resonance imaging study. *NeuroImage, 19*, 210–225.

Jang, K. E., Tak, S., Jung, J., Jang, J., Jeong, Y., & Ye, J. C. (2009). Wavelet minimum description length detrending for near-infrared spectroscopy. *Journal of Biomedical Optics, 14*(3), 34004. https://doi.org/10.1117/1.3127204

Jansma, J. M., Ramsey, N. F., Coppola, R., & Kahn, R. S. (2000). Specific versus nonspecific brain activity in a parametric N-back task. *NeuroImage, 12*, 688–697.

Jasdzewski, G., Strangman, G., Wagner, J., Kwong, K. K., Poldrack, R. A., & Boas, D. A. (2003). Differences in the hemodynamic response to event-related motor and visual paradigms as measured by near-infrared spectroscopy. *NeuroImage, 20*(1), 479–488.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation, *Electroencephalography Clinical Neurophysiology, 10*, 371-375.

Jöbsis, F. F. (1977). Non-invasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science, 198*, 1264–1267.

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., & Koeppe, R. A. (1997). Verbal Working Memory Load Affects Regional Brain Activation as Measured by PET. *Journal of Cognitive Neuroscience, 9*, 462–475.

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*, 637–671.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 33*, 615–622.

Kato, T., Kamei, A., Takashima, S., & Ozaki, T. (1993). Human visual cortical function during photic stimulation monitoring by means of near-infrared spectroscopy. *Journal of Cerebral Blood Flow & Metabolism, 13*, 516–520.

Kim, J.-J., Kim, M. S., Lee, J. S., Lee, D. S., Lee, M. C., & Kwon, J. S. (2002). Dissociation of working memory processing associated with native and second languages: PET investigation. *NeuroImage, 15*, 879–891.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55*, 352–358.

Klatzky, R. L., Giudice, N. A., Marston, J. R., Tietz, J., Golledge, R. G., & Loomis, J. M. (2008). An n-back task using vibrotactile stimulation with comparison to an auditory analogue. *Behavior Research Methods, 40*, 367–372.

Kleinschmidt, A., Obrig, H., Requardt, M., Merboldt, K.-D., Dirnagl, U., Villringer, A., & Frahm, J. (1996). Simultaneous Recording of Cerebral Blood Oxygenation Changes During Human Brain Activation by Magnetic Resonance Imaging and Near-Infrared Spectroscopy. *Journal of Cerebral Blood Flow & Metabolism, 16*, 817–826.

Koch, K., Pauly, K., Kellermann, T., Seiferth, N. Y., Reske, M., Backes, V., Stocker, T., Shah, J.N., Amunts, K., Kircher, T., Schneider, F., & Habel, U. (2007). Gender differences in the cognitive control of emotion: An fMRI study. *Neuropsychologia*, *45*(12), 2744-2754.

Kolb, B., & Whishaw, I, Q. (2009). The frontal lobes. In C. Linsmeier (Ed.), *Fundamentals of human neuropsychology* (6th ed., pp. 429-437). New York, NY: Worth Publisher.

Kreft, I., & De Leeuw, J. (1998). Introducing Multilevel Modeling. *Methods , 94*.

Kruggel, F., & von Cramon, D. Y. (1999). Temporal properties of the hemodynamic response in functional MRI. *Human Brain Mapping*, *8*, 259–271.

Kuznetsova, A., Brockhoff, P. B.,& Christensen, R. H. B (2015). lmerTest:Tests in Linear Mixed Effects Models. R package version 2.0-25. Available at: http://CRAN.R-project.org/package=lmerTest.

Kwok, O. M., Underhill, A. T., Berry, J. W., Luo, W., Elliott, T. R., & Yoon, M. (2008). Analyzing longitudinal data with multilevel models: an example with individuals living with lower extremity intra-articular fractures. *Rehabilitation Psychology*. *53*, 370–386.

León-Carrión, J., & León-Domínguez, U. (2012). Functional near-infrared spectroscopy (*f*NIRS): Principles and neuroscientific applications. In P. Bright (Ed.), *Neuroimaging–Methods* (pp. 47-74). Croatia: InTech.

Li, L., Zeng, L., Lin, Z.-J., Cazzell, M., & Liu, H. (2015). Tutorial on use of intraclass correlation coefficients for assessing intertest reliability and its application in functional near-infrared spectroscopy-based brain imaging. *Journal of Biomedical Optics*, *20*(5), 50801.

Li, T., Luo, Q., & Gong, H. (2010). Gender-specific hemodynamics in prefrontal cortex during a verbal working memory task by near-infrared spectroscopy. *Behavioural Brain Research*, *209*(1), 148–153.

Lin, S. Y. (2012). *f*MRI of a visual-patterns n-back working memory task in typical development. *Ms Thesis*, Institute of Medical Science University of Toronto.

Lina, J. M., Matteau-Pelletier, C., Dehaes, M., Desjardins, M., & Lesage, F. (2010). Wavelet-based estimation of the hemodynamic responses in diffuse optical imaging. *Medical Image Analysis*, *14*(4), 606–616.

Lindauer, U. (2006). Cerebral microvascular blood oxygenation and neuronal activity. In Shepro, D. (Ed.), *Microvascular Research*. Elsevier.

Lissitz, R. W. (2009). The Concept of Validity: Revisions, New Directions, and Applications. Information Age Publishing Inc.

Littell, R. C., Pendergast, J., & Natarajan, R. (2000). Tutorial in biostatistics - modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine*, *19*, 1793–1819.

Lloyd-Fox, S., Blasi, A., & Elwell, C. (2010). Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neuroscience &Biobehavioral. Reviews*, *34*, 269–284.

Logothetis, N. K. (2008). What we can do and what we cannot do with *f*MRI. *Nature*, *12*, 869–878.

Longford, N. T. (1993a). Random Coefficients Model. New York: Oxford University.

Matteau-Pelletier, C., Dehaes, M., Lesage, F., & Lina, J. M. (2009). 1/f noise in diffuse optical imaging and wavelet-based response estimation. *IEEE Transactions on Medical Imaging*, *28*(3), 415–422.

Max, L., & Onghena, P. (1999). Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research. *Journal of Speech, Language, and Hearing Research*, *42*(2), 261–270.

Mehagnoul-Schipper, D. J., van der Kallen, B. F. W., Colier, W. N. J. M., van der Sluijs, M. C., van Erning, L. J. T. O., Thijssen, H. O. M., … Jansen, R. W. M. M. (2002). Simultaneous measurements of cerebral oxygenation changes during brain activation by near-infrared spectroscopy and functional magnetic resonance imaging in healthy young and elderly subjects. *Human Brain Mapping*, *16*, 14–23.

Merzagora, A. C. (2010). Functional near infrared spectroscopy in the multimodal assessment of working memory impairments following traumatic brain injury, *PhD Thesis*, Drexel University, Philadelphia, PA.

Messick, S. (1989). Validity. In *Educational Measurement* (p. 13–105).

Miller, K. M., Price, C. C., Okun, M. S., Montijo, H., & Bowers, D. (2009). Is the N-back task a valid neuropsychological measure for assessing working memory? *Archives of Clinical Neuropsychology*, *24*(7), 711–717.

Molavi, B., & Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiological Measurement*, *33*(2), 259–70. https://doi.org/10.1088/0967-3334/33/2/259

Molteni, E., Contini, D., Caffini, M., Baselli, G., Spinelli, L., Cubeddu, R., … Torricelli, A. (2012). Load-dependent brain activation assessed by time-domain functional near-infrared spectroscopy during a working memory task with graded levels of difficulty. *Journal of Biomedical Optics*, *17*, 056005.

Muehlemann, T., Haensse, D., & Wolf, M. (2008). Wireless miniaturized in-vivo near infrared imaging. *Optics Express*, *16*, 10323–10330.

Murata, Y., Sakatani, Y., Katayama, Y., Fujiwara, N., Fukaya, C., & Yamamoto, T (2004). Decreases of blood oxygenation level-dependent signal in the activated motor cortex during functional recovery after resection of a glioma. *American Journal of Neuroradiology, 25,* 1242–1246.

Murphy, K., & Davidshofer, C. (2004). Psychological Testing: Principles and Applications. Reliability: The Consistency of Test Scores (6rd ed., pp. 116–129). Pearson

Niu, H., Li, Z., Liao, X., Wang, J., Zhao, T., Shu, N.,Zhao, X., & He, Y. (2013). Test-retest reliability of graph metrics in functional brain networks: a resting-state *f*NIRS study. *PloS One*, *8*, e72425.

Nystrom, L. E., Braver, T. S., Sabb, F. W., Delgado, M. R., Noll, D .C., & Cohen, J. D. (2000). Working memory for letters, shapes, and locations: *f*MRI evidence against stimulus-based regional organization in human prefrontal cortex. *NeuroImage*, *11*, 424–446.

Obrig, H., & Villringer, A. (2003). Beyond the visible-imaging the human brain with light. *Journal of Cerebral Blood Flow & Metabolism*, *23*, 1–18.

Obrig, H., Hirth, C., Junge-Hülsing, J. G., Döge, C., Wolf, T., Dirnagl, U., & Villringer, A. (1996). Cerebral oxygenation changes in response to motor stimulation. *Journal of Applied Physiology, 81*, 1174–1183.

Obrig, H., Wenzel, R., Kohl, M., Horst, S., Wobst, P., Steinbrink, J., …Villringer A. (2000). Near-infrared spectroscopy: Does it function in functional activation studies of the adult brain? *International Journal of Psychophysiology, 35*, 125–142.

Ogawa, S., Menon, R.S., Tank, D., Kim, S.-G., Merkle, K., Ellermann, J.M., & Ugurbil, K. (1993a) Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging: A comparison of signal characteristics with a biophysical model. *Biophysical Journal, 64*, 803–812.

Ogawa, S., Tank, D., Menon, R., Ellermann, J.M., Kim, S.-G., Merkle, K., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping using MRI. *Proceedings of the National Academy of Sciences of the United States of America, 89*, 5951–5955.

Okamoto, M., Dan, H., Sakamoto, K., Takeo, K., Shimizu, K., Kohno, S., … Dan, I. (2004). Three-dimensional probabilistic anatomical cranio-cerebral correlation via the international 10-20 system oriented for transcranial functional brain mapping. *NeuroImage, 21*, 99–111.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9*, 97-113.

Onyango, N. O. (2009). On the linear mixed effects regression (lmer) r function for nested animal breeding data. *CSBIGS, 4*(1):44–58.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*, 46-59.

Park, S., & Lake, E. T. (2005). Multilevel modeling of a clustered continuous outcome: nurses' work hours and burnout. *Nursing Research, 54*(6), 406–413.

Pellicer, A., & Bravo, M. D. C. (2011). Near-infrared spectroscopy: A methodology-focused review. *Seminars in Fetal and Neonatal Medicine, 16*, 42–49.

Plichta, M. M., Herrmann, M. J., Baehne, C. G., Ehlis, A. C., Richter, M. M., Pauli, P., & Fallgatter, A. J. (2006). Event-related functional near-infrared spectroscopy (*f*NIRS): Are the measurements reliable? *NeuroImage*, *31*, 116–124.

Plichta, M. M., Schwarz, A. J., Grimm, O., Morgen, K., Mier, D., Hadda, L., & Meyer-Lindenberg. (2012). A Test-retest reliability of evoked BOLD signals from a cognitive-emotive *f*MRI test battery. *NeuroImage*. *60*(3), 1746–1758.

Quaresima, V., Bisconti, S., & Ferrari, M. (2012). A brief review on the use of functional near-infrared spectroscopy (*f*NIRS) for language imaging studies in human newborns and adults. *Brain and Language*, *121*, 79–89.

Quené, H., & Van Den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, *43*, 103–121.

Ragland, J. D., Turetsky, B. I., Gur, R. C., Gunning-Dixon, F., Turner, T., Schroeder, L., … Gur, R. E. (2002). Working memory for complex figures: An *f*MRI comparison of letter and fractal n-back tasks. *Neuropsychology*, *16*, 370 – 379.

Rama, P., Martinkauppi, S., Linnankoski, I., Koivisto, J., Aronen, H., & Carlson, S. (2001): Working memory of identification of emotional vocal expressions: an *f*MRI study. *NeuroImage*, *13*, 1090–1101.

Reynolds, E. O. R., Wyatt, J. S., Azzopardi, D., Delpy, D. T., Cady, E. B., Cope, M., & Wray, S. (1988). New non-invasive methods for assessing brain oxygenation and hemodynamics. *British Medical Bulletin*, *44*, 1052–1075.

Rodriguez-Jimenez, R., Avila, C., Garcia-Navarro, C., Bagney, A., Aragon, A. M. De, Ventura-Campos, N., … Palomo, T. (2009). Differential dorsolateral prefrontal cortex activation during a verbal n-back task according to sensory modality. *Behavioural Brain Research*, *205*, 299–302.

Ruocco, A. C., Merzagora, A. C., Allen, S. L., Schultheis, M. T., Tuscan, L.-A., Lake, T., … Chute, D. (2007). One-month test-retest reliability of *f*NIRS during verbal fluency in a subject with discrete lesion of the left frontal lobe. *Archives of Clinical Neuropsychology*, *22*, 848.

Sakatani, K., Chen, S., Lichty, W., Zuo, H., & Wang, Y. P. (1999). Cerebral blood oxygenation changes induced by auditory stimulation in newborn infants

measured by near infrared spectroscopy. *Early Human Development*, *55*, 229–236.

Sato, H., Takeuchi, T., & Sakai, K. L. (1999). Temporal cortex activation during speech recognition: an optical topography study. *Cognition*, *73*, 55–66.

Sato, H., Yahata, N., Funane, T., Takizawa, R., Katura, T., Atsumori, H., … Kasai, K. (2013). A NIRS-*f*MRI investigation of prefrontal cortex activity during a working memory task. *NeuroImage*, *83*, 158–173.

Schecklmann, M., Ehlis, A.-C., Plichta, M. M., & Fallgatter, A. J. (2008). Functional near infrared spectroscopy: A long-term reliable tool for measuring brain activity during verbal fluency. *NeuroImage*, *43*, 147-155.

Schmidt, H., Jogia, J., Fast, K., Christodoulou, T., Haldane, M., Kumari, V., & Frangou, S. (2009). No gender differences in brain activation during the N-back task: An *f*MRI study in healthy individuals. *Human Brain Mapping*, *30*(11), 3609–3615.

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., & Wolf, M. (2014). A review on continuous wave functional near- infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage*, *85*, 6–27.

Scholtes, V. A., Terwee, C. B., & Poolman, R. W. (2011). What makes a measurement instrument valid and reliable? *Injury*, *42*(3), 236–240.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Sela, I., Izzetoglu, M., Izzetoglu, K., & Onaral, B. (2012). A working memory deficit among dyslexic readers with no phonological impairment as measured using the n-back task: An *f*NIR study. *PLoS ONE*, *7*, e46527.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Son, I. Y., & Yazıcı, B. (2006). Near-infrared imaging and spectroscopy for brain activity monitoring. *Advances in Sensing with Security Applications*, 341–372.

Speck, O., Ernst, T., Braun, J., Koch, C., Miller, E., & Chang, L. (2000). Gender differences in the functional organization of the brain for working memory. *Neuroreport*, *11*, 2581–2585.

Strait, M., & Scheutz, M. (2014). What we can and cannot (yet) do with functional near infrared spectroscopy. *Frontiers in Neuroscience. 8*, 1-12.

Strangman, G., Boas, D. A, & Sutton, J. P. (2002). Non-invasive neuroimaging using near-infrared light. *Biological Psychiatry*, *52*, 679–693.

Strangman, G., Culver, J. P., Thompson, J. H., & Boas, D. A. (2002). A quantitative comparison of simultaneous BOLD *f*MRI and NIRS recordings during functional brain activation. *NeuroImage*, *17*, 719–731.

Strangman, G., Goldstein, R., Rauch, S. L., & Stein, J. (2006). Near-infrared spectroscopy and imaging for investigating stroke rehabilitation: test-retest reliability and review of the literature. *Archives of Physical Medicine and Rehabilitation*, *87*, 12–19.

Stuss, D., & Alexander, M. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research*, *63*, 289–298.

Tisdall, M. M. (2007). Non-invasive near infrared spectroscopy: A tool for measuring cerebral oxygenation and metabolism in patients with traumatic brain injury. *Phd Thesis*. University of London.

Toronov, V., Walker, S., Gupta, R., Choi, J. H., Gratton, E., Hueber, D., & Webb, A. (2003). The roles of changes in deoxyhemoglobin concentration and regional cerebral blood volume in the *f*MRI BOLD signal. *NeuroImage*, *19*, 1521–1531.

Toronov, V., Webb, A., Choi, J. H., Wolf, M., Michalos, A., Gratton, E., & Hueber, D. (2001a). Investigation of human brain hemodynamics by simultaneousnear-infrared spectroscopy and functional magnetic resonance imaging. *Medical Physics*, *28*, 521–527.

Toronov, V., Zhang, X., & Webb, A. G. (2007). A spatial and temporal comparison of hemodynamic signals measured using optical and functional magnetic resonance imaging during activation in the human primary visual cortex. *NeuroImage*, *34*, 1136–1148.

Totaro, R., Barattelli, G., Quaresima, V., Carolei, A., & Ferrari, M. (1998). Evaluation of potential factors affecting the measurement of cerebrovascular reactivity by near-infrared spectroscopy. *Clinical Science, 95*, 497–504.

Veltman, D. J., Rombouts, S. A., & Dolan, R. J. (2003). Maintenance versus manipulation in verbal working memory revisited: an *f*MRI study. *NeuroImage, 18*, 247–256.

Villringer, A., & Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences, 20*, 435–442.

Villringer, A., Planck, J., Hock, C., Schleinkofer, L., & Dirnagl, U. (1993). Near-infrared spectroscopy (NIRS): A new tool to study hemodynamic changes during activation of brain function in human adults. *Neuroscience Letters, 154*, 101–104.

Watanabe, A., Matsuo, K., Kato, N., & Kato, T. (2003). Cerebrovascular response to cognitive tasks and hyperventilation measured by multi-channel near-infrared spectroscopy. *The Journal of Neuropsychiatry and Clinical Neurosciences, 15*, 442–449.

Weissbluth, M. (1974). Hemoglobin. In A. Kleinzeller, G. F. Springer & H. G. Wittmann (Eds.), *Hemoglobin: Cooperativity and electronic properties (Molecular biology, biochemistry and biophysics, Vol 15)* (pp. 10-15). Heidelberg, Berlin: Springer-Verlag.

Whittemore, S. (2004). Oxygen transport: The role of hemoglobin. *The*

Widmaier, E. P., Raff, H., & Strang, K. T. (2013). Respiratory physiology. *Vander's Human Physiology: The mechanisms of Body Function* (13th ed., pp. 467-473).

Wiggins, I. M., Anderson, C. A., Kitterick, P. T., & Hartley, D. E. H. (2016). Speech-evoked activation in adult temporal cortex measured using functional near-infrared spectroscopy (fNIRS): Are the measurements reliable? *Hearing Research, 339*, 142–154. https://doi.org/10.1016/j.heares.2016.07.007

Wolf, M., Wolf, U., Choi, J. H., Gupta, R., Safonova, L. P., Paunescu, L. A., … Gratton, E. (2002). Functional frequency-domain near-infrared spectroscopy detects fast neuronal signal in the motor cortex. *NeuroImage, 17*, 1868–1875.

Wyatt, J. S., Cope, M., Delpy, D. T., Wray, S., & Reynolds, E. O. (1986). Quantification of cerebral oxygenation and haemodynamics in sick newborn infants by near-infrared spectrophotometry. *Lancet*, *2*, 1063–1066.

Zabel, A., & Chute, D. L. (2002). Near infrared spectroscopy in functional brain imaging. *Journal of Head Trauma Rehabilitation*, *17*, 477–488.

Zald, D. H., & Curtis, C. (2006). Brain imaging and related methods, In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 173–187). American Psychological Association.

Zhang, Q., Brown, E. N., & Strangman, G. E. (2007). Adaptive filtering for global interference cancellation and real-time recovery of evoked brain activity: a Monte Carlo simulation study. *Journal of Biomedical Optics*, *12*(4), 44014.

Zhang, X. F., Toronov, V. Y., & Webb, A. G. (2005a). Simultaneous integrated diffuse optical tomography and functional magnetic resonance imaging of the human brain. *Optics Express*, *13*, 5513–5521.

Zhang, X. F., Toronov, V. Y., & Webb, A. G. (2006). Integrated measurement system for simultaneous functional magnetic resonance imaging and diffuse optical tomography in human brain mapping. *Review of Scientific Instruments*, *77*, 114301.

Zhang, Y., Brooks, D. H., Franceschini, M. A., & Boas, D. a. (2005). Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging. *Journal of Biomedical Optics*, *10*(1), 11014.

Zurowski, B., Gostomzyk, J., Grön, G., Weller, R., Schirrmeister, H., Neumeier, B., … Walter, H. (2002). Dissociating a common working memory network from different neural substrates of phonological and spatial stimulus processing. *NeuroImage*, *15*, 45–57.

Appendix A

KATILIMCI BİLGİLENDİRME FORMU

Bu çalışmanın amacı, laboratuvar koşullarında beynin yürütücü işlevlerini incelemek üzere kullanılan çalışma belleği testinin uygulanması sırasında elde edilecek beyin hemodinamik aktivitesinin incelenmesidir.

Çalışmaya başlamadan önce araştırmacı sizden bir takım soruları yanıtlamanızı isteyecek ve ardından sizi deneysel oturumların gerçekleşeceği odaya alacaktır. Bu odadaki koltukta rahat bir pozisyonda oturmanız, uygulamalar boyunca konuşmamanız ve pozisyonunuzu korumanız çalışmanın başarısı açısından oldukça önemlidir. Araştırmacı ön beyin aktivitesini takip etmek amacıyla alın bölgenize bir bant yerleştirecektir. Bu işlem bittikten sonra araştırmacı yanınızdan ayrılıp yan odada sizi bekliyor olacaktır. Çalışma boyunca, ekrandan sunulan yönergeleri dikkatlice okumanız ve sizden istenenleri olabildiğince doğru bir biçimde yerine getirmeniz gerekmektedir.

Çalışma kapsamında katılımcılardan elde edilen veriler isim kullanılmaksızın analizlere dâhil edilecektir. Çalışma başında size bir katılımcı numarası verilecek ve isminiz araştırma raporunda yer almayacaktır.

Katılımınız araştırma hipotezinin test edilmesi ve yukarıda açıklanan amaçlar doğrultusunda literatüre sağlayacağı katkılar bakımından oldukça önemlidir. Ayrıca katılımınızın psikoloji alanın gelişmesi açısından da bir takım faydaları bulunmaktadır.

Çalışmaya katılmanız tamamen kendi isteğinize bağlıdır. Katılımı reddetme ya da çalışma sürecinde herhangi bir zaman diliminde devam etmeme hakkına sahipsiniz. Eğer görüşme esnasında katılımınıza ilişkin herhangi bir sorunuz olursa, araştırmacıyla iletişime geçebilirsiniz.

Okudum, kabul ediyorum.

Appendix B

KATILIMCI İZİN FORMU

Çalışmanın amacını ve içeriğini ..... katılımcı numarasına sahip katılımcıya açıklamış bulunmaktayım. Çalışma kapsamında yapılacak işlemler hakkında katılımcının herhangi bir sorusu olup olmadığını sordum ve katılımcı tarafından yöneltilen bütün soruları yanıtladım.

Tarih: Araştırmacının İmzası:

..... / ..... / ....

Çalışmanın amacı ve içeriği hakkında açıklamaların yer aldığı "Katılımcı Bilgilendirme Formu"nu okudum. Araştırmacı çalışma kapsamındaki haklarımı ve sorumluluklarımı açıkladı ve kendisine yönelttiğim bütün soruları açık bir şekilde yanıtladı. Sonuç olarak, uygulama esnasında şahsımdan toplanan verilerin bilimsel amaçlarla kullanılmasına izin verdiğimi ve çalışmaya gönüllü olarak katıldığımı beyan ederim.

Tarih: Katılımcının İmzası:

..... / ..... / ....

Appendix C

KATILIMCI BİLGİ FORMU

AD-SOYAD:	TELEFON NUMARASI:
CİNSİYET:	e-MAIL:
YAŞ:	OKUL:
MESLEK:	

TA = SpO_2 = PR =

Aşağıdaki sorulyu yanıtlarken size en uygun olan numarayı yuvarlak içine alınız. (0= hiç yorgun değil, 7= çok yorgun)

1. Şu anda kendinizi ne kadar yorgun hissediyorsunuz?

0 -----1 -----2 -----3 -----4 -----5 -----6 -----7

Aşağıdaki soruları yanıtlarken lütfen durumunuzu en iyi yansıtan seçeneği işaretleyiniz.

1. Yakın zamanda (son 1 sene dahil) başka bir psikoloji deneyine katıldınız mı?

○ Evet ○ Hayır

Yanıtınız "Evet" ise 2.sorudan, "Hayır" ise 3. sorudan devam ediniz.

2. Hangi deneye katıldınız?

…….

3. Herhangi ciddi bir görme bozukluğunuz var mı?

○ Evet ○ Hayır

4. Herhangi bir psikolojik rahatsızlık geçmişiniz var mı?

○ Evet ○ Hayır

Yanıtınız "Evet" ise 5. sorudan, "Hayır" ise 7. sorudan devam ediniz.

5. Bir ruh sağlığı çalışanı tarafından rahatsızlığınıza konulan tanı nedir?

……

6. Rahatsızlığınız ile ilgili kullandığınız ilaç(lar) var mı?

○ Evet,……isimli ilaç(lar)ı kullandım/kullanmaktayım.

○ Hayır

7. Herhangi bir nörolojik hastalık geçmişiniz var mı?

○ Evet ○ Hayır

Yanıtınız "Evet" ise 8. sorudan, "Hayır" ise 10. sorudan devam ediniz.

8. Bir uzman tarafından hastalığınıza konulan tanı nedir?

……

Hastalığınız ile ilgili kullandığınız ilaç(lar) var mı?

○ Evet,……isimli ilaç(lar)ı kullandım/kullanmaktayım.

○ Hayır

9. Daha önce kafa travması geçirdiniz mi?

○ Evet ○ Hayır

10. Düzenli olarak kullandığınız ilaç(lar) var mı?

○ Evet ○ Hayır

Yanıtınız "Evet" ise 12. sorudan, "Hayır" ise 13. sorudan devam ediniz.

11. Lütfen kullandığınız ilaç(lar)ı ve ilaç(lar)ın kullanım amaçlarını belirtiniz.

İlaç(lar):……..……..…… Kullanım amacı:……..……..……

12. Herhangi bir kalp rahatsızlığı tanısı aldınız mı?

○ Evet ○ Hayır

Yanıtınız "Evet" ise 14. sorudan, "Hayır" ise 15. sorudan devam ediniz.

13. Size konulan tanıyı belirtiniz.

……

14. Herhangi bir ameliyat/ operasyon geçirdiniz mi?

○ Evet ○ Hayır

Yanıtınız "Evet" ise 16. sorudan, "Hayır" ise 17. sorudan devam ediniz.

15. Geçirdiğiniz ameliyatı/operasyonu lütfen belirtiniz.

Ameliyat/operasyon:…… Ameliyat/operasyon tarihi:……

16. Dün akşam kaç saat uyudunuz?

○ 5 saatten az ○ 6-8 saat ○ 8 saatten fazla

17. Aneminiz (kansızlığınız) var mı?

○ Evet ○ Hayır

Appendix C (continued)

Lütfen ilk sütunda sıralanmış olan aktiviteleri yaparken veya söz konusu aletleri kullanırken hangi elinizi tercih ettiğinizi ilgili sütunda işaretleyiniz.

Edinburgh El Tercihi Envanteri					
	Her zaman sol	Genelde sol	Tercihim yok	Genelde sağ	Her zaman sağ
Yazma					
Fırlatma					
Makas					
Diş fırçası					
Bıçak					
Kaşık					
Kibrit					
Mouse					