Learning to Relax and Attend: Investigating Methods to Analyze Neurofeedback Data from Nepalese Children's Mind-Full Sessions

by

Srilekha Kirshnamachari Sridharan

B.E., Anna University, 2012

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
School of Interactive Arts and Technology
Faculty of Communication, Art and Technology

SIMON FRASER UNIVERSITY

Summer 2017

Approval

Name:	**Srilekha Kirshnamachari Sridharan**
Degree:	**Master of Science**
Title:	***Learning to Relax and Attend: Investigating Methods to Analyze Neurofeedback Data from Nepalese Children's Mind-Full Sessions***
Examining Committee:	**Chair:** Dr. Kate Hennessy Associate Professor

Dr. Alissa Antle
Senior Supervisor
Professor

Dr. Bernhard Riecke
Supervisor
Associate Professor

Dr. Steve DiPaola
External Examiner
Professor
School of Interactive Arts + Technology
Simon Fraser University

Date Defended/Approved: 10th August 2017

Abstract

Mind-Full (Nepal) consists of three neurofeedback (NF) games designed to help young children living in extreme poverty learn to self-regulate relaxation and attention. In this thesis, I present the methodological process used to analyze the Mind-Full's log data that was collected from a field-study conducted in Nepal (Antle et al., 2015). The results of this analysis showed that there was no significant improvement in relaxation, attention and game performance of the children across sessions in all three games. There was no correlation between the dependent measures derived from headset generated relaxation/attention indices and brainwave amplitudes. I discuss reasons for these findings, grounded in the previous NF studies. Based on my results and previous works, I recommend approaches to data analysis methods for future NF studies including how to pre-process data, choose dependent measures and sample sessions for across sessions analysis.

Keywords: Relaxation; Attention; EEG; Neurofeedback; Children; Quantitative Data Analysis.

Dedication

I dedicate this thesis to my grandparents and my pet Shaggy whose kindness and selfless acts will keep inspiring me for life.

Acknowledgements

I thank my professor, Dr. Alissa Antle, for her guidance and supervision, without whom this thesis would not be possible. She taught me to carry out research, how to approach the research problems with confidence and helped with my thesis in choosing the right path. I am also grateful to her for motivation and guidance that she has provided to me ever since I came to SFU.

I thank my committee, Dr. Bernhard Riecke, who gave valuable input for this thesis. He also taught me how to design controlled experiment and his course on quantitative methodologies gave me confidence in statistical analysis, especially when I was a newbie to stats.

I thank Emily and Sairaam, who reviewed my quantitative methods and report, addressed my statistical questions, and provided useful feedbacks on my analysis. I thank Siddharth and Annique-Elise for proof-reading my thesis. I thank Aishwarya for proof-reading my statistical reports. I thank my Wearable Therapeutics team member Jacob, who assisted me in running the headset validation study. I thank my SFU colleagues and friends who have been a good support network and have guided me at both personal and professional level.

I thank Siddharth for his moral support. I thank my parents and my brother who have been my guides and greatest pillars of support.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ADHD	Attention Deficit Hyperactive Disorder
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
BCI	Brain Computer Interaction
DV	Dependent Variable
EC	Eyes Closed
EEG	Electroencephalogram
EEG NF	Electroencephalogram-based Neurofeedback
EO	Eyes Open
FASD	Fetal Alcohol Spectrum Disorder
FFT	Fast Fourier Transform
GQSP	Good Quality Signal Percentage
HA	High Alpha
HB	High Beta
Hz	Hertz
IQR	Inter Quartile Range
IV	Independent Variable
LA	Low Alpha
LB	Low Beta
M	Mean
MANOVA	Multi-variate Analysis of Variance
NF	Neurofeedback
NGO	Non-Governmental Organisation
NMM	Neurosky MindWave Mobile
PS	Poor Signal
PT	Percentage Time
PT-A	Percentage of time spent above the attention index 40
PT-R	Percentage of time spent above the relaxation index 40
R/A	Relaxation and Attention
REM	Rapid Eye Movement
RM	Repeated Measurement

SCP	Slow Cortical Potential
SD	Standard Deviations
SFU	Simon Fraser University
SMR	Sensorimotor Rhythm
T/A	Theta / Alpha
T/B	Theta / Beta
TT	Time taken to collect five tokens

Glossary

Amplitude	Strength or Voltage of a signal, measured in volts.
Attention	Refers to the attention score given by the headset Neurosky MindWave Mobile (NMM). It ranges from 0 to 100 (no units). Value below 40 refers non-attentive state. It is sampled at 1 Hz.
Attention Deficit Hyperactive Disorder	A psychological phenomenon in which the person is unable to pay attention and control behaviour in an age-appropriate way.
Baseline	In NF studies, baseline measures generally refer to the data collected (e.g., relaxation index, EEG amplitudes) in idle condition, i.e., when no task was performed by the user.
Brain-Computer Interface	Refers to modern headset that reads the EEG signals from the brain and send it to a computer.
Brain waves	Refers to the oscillations of electrical impulses, measured using electrodes in the scalp. Unit is Hertz (Hz). Generally, the range of brainwaves measured is from 1–50 Hz. The different brainwaves are alpha (8 to 13 Hz), beta (12 to 30 Hz), theta (4 to 8 Hz), gamma (30 to 50 Hz), and delta (1 to 4 Hz).
Continuous sessions	In the context of session sampling in the across-session analysis, if most or all sessions are considered for the analysis, then the sampling method is termed as continuous session sampling. For example, in a 25-sessions study, all sessions from sessions 1 to 24 were considered for the data analysis.
Data pre-processing	Refers to the steps involved in transformation of raw data to a data set that is ready for analysis. This includes dependent measure calculation, noise exclusion, outlier exclusion, and missing data substitution.
Default time	Represents the default time (in seconds) set in the Mind-Full game, for which the player should hold their relaxation or attention score above a certain value to get a game token.
Discrete sessions	In the context of session sampling in the across-session analysis, if only a selected few sessions are considered for the analysis, then such a sampling method is termed as discrete session sampling. For example, in a 25-sessions study, only sessions 2 and 20, or only the initial sessions 2, 3, 4 and the last sessions 23, 24, 25 are considered for analysis.
Electroencephalogram	Refers to a non-invasive technique used to measure the electric signals generated in the cerebral cortex of the brain.

Electrodes	A metal conductor that is placed on the scalp to catch electrical signals of the brain.
Frequency	Number of waves per second. In this thesis, frequency refers to the brainwave frequency.
Good Quality Signal Percentage	Term used in this thesis to represent the percentage of time in a session in which the signal quality was good. It is calculated as [(no. of seconds in a session in which the signal quality was good / total no. of seconds in a session) * 100].
Group	In this thesis, group refers to the participant groups G1 and G2. G1 had nine children as participants and G2 had 12 children as participants at the start of the study.
Mean attention	Mean value of the attention score given by the NMM per game per session per participant.
Mean high alpha amplitude	Refers to the average amplitude of high alpha (10 to 11.75 Hz) per session per game per participant. The amplitude value is given by NMM at the rate of 1 Hz.
Mean low alpha amplitude	Refers to the average amplitude of low alpha (7.5 to 9.25 Hz) per session per game per participant. The amplitude value is given by NMM at the rate of 1 Hz.
Mean low beta amplitude	Refers to the average amplitude of low beta (13 to 16.75 Hz) per session per game per participant. The amplitude value is given by NMM at the rate of 1 Hz.
Mean relaxation	Mean value of the relaxation score given by the NMM per game per session per participant.
Mean theta amplitude	Refers to the average amplitude of theta (3.5 to 6.75 Hz) per session per game per participant. The amplitude value is given by NMM at the rate of 1 Hz.
Min_time	In the Mind-Full game, min_time or minimum time (in seconds) represents the pre-set duration of time for which the player is expected to hold their relaxation/attention value above a certain value (i.e. Threshold) to get a game token.
Mind-Full	The series of neurofeedback game that was used in the Nepal study to self-regulate the calmness and attention of the children. This thesis analyzes the data collected from this game.
Muse	Consumer-grade EEG headset available in the market. This headset has four electrodes.
Neurofeedback	Technique in which real-time brain activity is passed as feedback in a form of stimuli (e.g., visual or audio) to improve the mental state (emotional or cognitive) of the user such as relaxation, attention, etc.

Neurosky MindWave Mobile	Consumer-grade EEG headset available in the market. This headset has one electrode.
Noise	In Neurofeedback studies, noise represents any drastic or outlying data point in the raw signal. In Neurosky MindWave Mobile (NMM), noise is indicated by the attribute POOR_SIGNAL > 0.
Paraglider game	A type of NF game in the Mind-Full game series that aims to self-regulate the relaxation of the children. The player would be able to land the paraglider when they are able to hold the relaxation value above a certain level of pre-set relaxation value over a pre-set amount of time. This action rewards the player with a game point.
Percentage time	The percentage of time in a session that was spent above the relaxation (or attention) index of 40
Pinwheel game	A type of NF game in the Mind-Full game series that aims to self-regulate the relaxation of the children. This is more of a warm-up game. The player can spin the pinwheel when they are able to hold the relaxation value above a certain level of pre-set relaxation value over a pre-set amount of time. This action rewards the player with a game point.
Pre-frontal cortex	The region of brain that is close to the forehead.
Relaxation	Term referring relaxation score given by the headset Neurosky MindWave Mobile. It ranges from 0 to 100 (no. units). A value below 40 refers to a stressed state of mind. It is sampled at 1 Hz.
Session	Represents an experimental session. In the context of the current thesis, it represents a Mind-Full game session where a child has played all three games of Mind-Full.
Session sampling	The method of selecting certain sessions for data analysis.
Stones game	A type of NF game in the Mind-Full game series that aims to self-regulate the attention of the children. The player can put the stone into a basket when they are able to hold the attention value above a certain level of pre-set attention value over a pre-set amount of time. Putting five stones into the basket rewards a player with a game point.
Threshold	In the Mind-Full game, threshold represents the pre-set value of relaxation or attention. The player should be able to hold their relaxation or attention value above this pre-set value to get a game point or token.
Time taken to get tokens in a session	Attribute referring to the dependent variable of a research question in this current thesis. The total number of seconds taken to collect all game points per session per game = time taken to get tokens in a session.

Token	Represents the game points or reward points in the Mind-Full game.

Chapter 1. Introduction

1.1. Research Problem

According to the World Bank, 400 million children live in extreme poverty[1]. Children's exposure to extreme poverty increases the risk factors for trauma and negatively impacts their cognitive development and behaviour (Walker et al., 2007). Risk factors include domestic violence and maltreatment of children by their parents. These risk factors can induce trauma that can cause issues such as anxiety disorders, attentional issues, and poor academic performance in school (Bellis, 2001). Children studying in the Nepal House Kaski school, in Pokhara, live in extreme poverty. The children were reported to have behavioural issues like hyperactivity, inattentiveness, emotional imbalance, poor hygiene, aggressiveness, and unruly behaviour in the classroom and playground. The children's issues prevent them from concentrating on their education because they are agitated in class and engage in negative activities in the classroom and playground. The negative behaviours of the children could be improved by training, which will help them overcome their emotional, behavioural, and attentional issues (Diamond & Lee, 2011). To date, western-trained Nepali counselors in the school have worked with the children on breathing, yoga, play therapy and other counseling activities. Success has been slow and intermittent. To address this problem, Dr. Antle developed a neurofeedback (NF) game called "Mind-Full" to be used as a tool by the counselors in the Nepal House Society. This game may help the children to learn and practice self-regulation of anxiety and attention, with the hope that this learning would transfer to other contexts such as the classroom and playground. With this game, an equivalent-group experimental field study was conducted by Dr. Antle with 21 children. In this study, approximately half the number of children underwent counselor-facilitated sessions with the Mind-Full game over 6–7 weeks while the other group served as a waitlist control (Antle, Chesick, Levisohn, Sridharan, &

[1] http://www.worldbank.org/en/news/press-release/2013/10/10/report-finds-400-million-children-living-extreme-poverty

Tan, 2015). For ethical reasons, the other group subsequently was provided with the same treatment. Three behavioural assessments were administered by the counsellors to all children as a part of this study. The first assessment was conducted before the start of the Mind-Full training. The second assessment was conducted after the first group had their Mind-Full training and the final one was conducted after the second group had their Mind-Full training. During the treatment, the game data and electro-encephalogram (EEG) headset data were collected as log files. Preliminary analysis of the behavioural assessment data comparing the first and second assessment, showed significant improvement in the children's ability to calm down and pay attention in a variety of contexts (e.g., the classroom and playground) (Antle et al., 2015). However, this type of data is inherently biased since the counselors and teachers who assessed the children, working with a trained facilitator, were not blind to the condition.

Further analysis of the game log data may add to understanding if the Mind-Full training treatment helped the children learn self-regulation of anxiety and attention, which could complement the behavioural improvement. Within-subject analysis of log data may show patterns of game performance and electroencephalogram (EEG) data across the sessions in each treatment group. In this thesis, I investigate if there is a general improvement in the pattern of dependent measures across the sessions, as children get better with time in self-regulating their calmness/attention during Mind-Full gameplay. Previous work with healthy adults showed variation in learning across sessions, where some studies found a pattern (e.g., initial decrease, followed by a large gain, followed by plateau), and others found no pattern (Gruzelier, 2014). If there is a pattern of improvement, it is possible that maturation or learning could lead to a noticeable improvement. Conversely, trauma or other negative life events could contribute to a lack of improvement. Over time as trauma continues, children's abilities to self-regulate typically decreases for children living in poverty in contrast to typical child development (Armsworth & Holaday, 1993). I was interested in exploring the log data to look for patterns. We don't know yet how to analyze the game performance or EEG headset data collected from the Nepal study. As the data was collected from a field environment, there are several constraints to deal with such as inconsistent number of sessions, missing sessions, noisy data, connectivity issues, participant dropping out in the middle of the study, *etc*. Additionally, it was also unclear if I should analyze all the session data or sample them before analysis. Most of the previous researchers, exploring electroencephalogram-based neurofeedback (EEG NF) treatments for children, had analyzed the raw EEG signals and/or looked at subjective behavioural assessment data to understand the improvements (or lack thereof) in self-regulation (Fuchs, Birbaumer, Lutzenberger, Gruzelier, & Kaiser, 2003; Lubar, Swartwood, Swartwood, & O'Donnell, 1995). However, the Mind-Full game mechanics provide neurofeedback based on the relaxation and attention indices provided by the Neurosky MindWave Mobile (NMM). The raw brainwave data was not collected during the study, as it was not possible to collect it at a high enough frequency to be suitable for analysis because of storage limitations on the tablets and internet bandwidth issues that limited file size in uploading the log files. Therefore, I needed to determine how to analyze the log data I did have, which included: the Neurosky relaxation and attention indices, game performance data (e.g., tokens collected), and the EEG amplitudes collected from the NMM over the study period. Prior to this, I needed to determine how to handle constraints such as noise exclusion, outlier detection, session and participant considerations.

1.2. Summary of Previous Research

Several NF interventions that use consumer-grade headsets, such as NMM, have been developed for children living in industrialized countries to improve their focus and calmness. These research works indicated that children can use this type of BCI system easily. However, many of these studies were not methodologically rigorous (e.g., collected only subjective measures, lack of proper reasoning behind data analysis decisions) or had a different study design (only within-session comparison). The methods used could not be directly applied to the Mind-Full log data. For example, FOCUS is an application that intends to increase the engagement of children in textbook reading by using an augmented display over a textbook (Huang et al., 2014). For the training group, the EEG engagement index of the children triggers the training. Their results showed that the group that received NF between their reading sessions had a better mean EEG engagement index than the group that had a NF session after the reading sessions. However, the study did not collect data across sessions, and it could not substantiate if the intervention could improve self-regulation over time. In another study, NF using the Neurosky headset was used in traditional video games to help children with fetal alcohol spectrum disorder (FASD) to improve their engagement level. Biofeedback triggered a texture-based overlay that obfuscated the game when attentional levels were low (Mandryk et al., 2013). The study suggested that the children's engagement improved while playing the game and survey results revealed that the children enjoyed the game. However, the authors did not analyze the data across sessions, and they compared within-session changes even though they had data for 12 weeks. A BCI-based puzzle was built for children with attention-deficit hyperactive disorder (ADHD), with the aim to improve their attentive state by solving puzzles (Lim et al., 2010). Though there was no significant improvement reported in surveys by the parents and teachers, the work claims that the children accepted and underwent the BCI training with ease. Even though the study was carried out for 24 sessions, they only considered sessions conducted in week 0 and week 14 for analysis, and they did not look at the relaxation or attention indices provided by the NMM. Thomas et al. (2013) studied if NF training can improve children's attention by embedding the NF within a number game. Even though the attention indices were collected across sessions, the work did not investigate if there was any improvement in attention across sessions through inferential statistics. Some NF studies, conducted with adults using consumer-grade EEG, had proper statistical methodology, however, their study designs had only one session (e.g., Belkofer & Konopka, 2008; Lee, 2009; Stinson & Arthur, 2013).

Some studies have used NF training for ADHD children under clinical conditions with research-grade EEG headsets. Lubar et al. (1995) found that the NF mechanism was effective in self-regulating attention by conducting a study with complex EEG headsets on 19 children. A randomized between-subjects controlled experiment was conducted on children with auditory and visual NF by McDonald (1974). The findings showed that both groups had improvement in attention over time. Bakhshayesh et al. (2011) found that there was an improvement in behavioural performance of the children after NF training in the training group compared to the control group. Drechsler et al. (2007) observed an improvement in the training group on attention related EEG measures by conducting a clinical NF training with 30 children. Hillard et al. (2013) and Takahashi et al. (2014) investigated the effects of NF training on ADHD children under clinical conditions, and demonstrated improvement in attention measures over time. Even though these studies followed a rigorous methodology, all these studies used complex EEG systems, which read electrical signals from the scalp. These studies have neither received the relaxation and attention indices from a consumer-grade headset nor collected any other game performance metrics. Mind-Full data analysis could adapt certain data pre-processing strategies (such as inclusion and exclusion criteria for session and participants); however, for analyzing data from our Mind-Full Nepal study, the dependent measures and certain data-preprocessing (e.g., noise cancellation) criteria cannot be directly applied from these studies.

1.2.1. Need for Current Research

As seen in Section 1.2, the previous studies on the NF interventions developed for self-regulation of relaxation and attention over time either used a commercial grade EEG headset in a laboratory setting and did not have a methodology close to our Mind-Full study or involved rigorous evaluation conducted in a sophisticated environment using complex electrodes for measuring EEG data or. As per the author’s knowledge, there are no other methodological approaches in existing research that can be directly applied to our dataset to analyze patterns across sessions of children’s game performance and relaxation/attention data from commercial grade EEG headsets.

1.3. Goal and Contribution

The overall goal of this thesis is to determine and apply appropriate statistical analysis to the Nepal log data, including: Neurosky relaxation and attention index data; proprietary Neurosky EEG data; and Mind-Full game performance data. This analysis will investigate if there are any changes in brain activity and game performance across sessions. I might expect a gradual improvement in the children’s ability to relax and focus their attention, as captured by game performance and/or brain activity data, over the course of their training. The main contribution of this thesis is methodological; investigating and determining how to analyze this type of data in the Nepal study and other Mind-Full studies. A secondary contribution is practical; applying the analytical methods in order to help me determine if (and how) the Mind-Full treatment has helped the children to learn and practice self-regulation of anxiety and attention during their sessions. Additionally, a tertiary contribution from this thesis is the informal validation study of the NMM headset (used in Mind-Full). This study was conducted with adults as a part of my internship with Wearable Therapeutics.

1.3.1. Stakeholders

(1) Researchers who intend to analyze NF data from consumer-grade headsets in multi-sessional field studies. The methodological implications of the current thesis will inform researchers about the data analysis techniques for NF research like Mind-Full. This process includes selecting dependent measures, sampling sessions, handling data constraints (e.g., noise, outlier and missing data), performing appropriate statistical analysis, and other data-related precautions while designing/developing the intervention.

(2) Researchers who are planning to use consumer-grade EEG headset NMM as a research instrument in non-laboratory field environments and who want to know the validity and limitations of the headset.

(3) Consumer-grade EEG manufacturers who want to show evidence of the data quality and effectiveness of their proprietary algorithm in a non-laboratory environment where it is difficult to control all factors impacting the quality of the EEG data.

(4) School therapists and teachers who want to know if (and how) NF may benefit the children if they learn and practice self-regulation using a Mind-Full-like NF application with consumer-grade EEG headsets.

1.4. Scope

For this thesis, I set the scope as follows:

(1) Provide a literature survey of NF studies that have used both consumer-grade and research-grade EEG headsets.

(2) Develop measures and apply statistical methods to analyze the log data collected from the Mind-Full Nepal study. The scope of this thesis is to analyze change in the dependent measures across sessions. Other work in progress deals with investigating the change in the dependent measures between selected sessions and within-session change, and other alternate methodologies mentioned in the discussion of the thesis.

(3) Provide results to understand if there were patterns in children’s relaxation and attention across sessions. This analysis would help us in giving methodological

implications – that would be helpful in conducting future Mind-Full-like studies and analyzing their data.

(4) Provide evidence from an informal headset validation study conducted in a controlled environment with adults, where I tried to investigate if the NMM headset can distinguish relaxation, anxiety and attention measures. This study would inform future EEG NF studies about the effectiveness of the NMM headsets.

1.5. Research Questions

1.5.1. Headset Validation Study

Consumer-grade EEG devices have their own proprietary algorithms that calculate affective and cognitive indices (typically in interval scales, e.g., 0–100) such as attention, meditation, frustration, engagement, and memory from the brainwave frequency and amplitude. These values can be used as controllers in human–robot interaction (Vourvopoulos & Liarokapis, 2012; Vourvopoulos & Liarokapis, 2014), for NF in everyday gaming systems (Antle et al., 2015), and as attention regulators in learning systems (Huang et al., 2014). Use of consumer-grade EEG devices make these systems readily available, easy to use, and inexpensive (Huang et al., 2014). Hence, for the success of systems, it is important to know if the proprietary algorithms of consumer-grade EEG devices can reliably measure affective and cognitive states. As a part of my Mitacs Internship with Wearable Therapeutics Inc. Vancouver, I conducted a controlled lab experiment to test the effectiveness of a consumer-grade EEG headset, the Neurosky MindWave Mobile (NMM). The goal of the study was to inform the company if NMM could distinguish the relaxed state (from anxious state) and attentive state (from non-attentive state), and to provide them with a methodological approach to determine the time spent on relaxation and attention from the NMM headset. This approach will be used by the company for their future research to validate their relaxation and attention self-regulating intervention. The headset computes scores of relaxation and attention based on EEG signals. This study aimed to address the research questions: Is the proprietary algorithm of the NMM headset effective in detecting users' relaxation (when compared to a stressed state) and user's attentive state (or non-attentive state)?

I conducted the study with adults as participants. Even though the study does not directly let us generalize the results to children, it does inform us that the proprietary algorithm of the headset NMM can identify relaxed, stressed and attentive state of adults, which may extend to children who have similar brainwave-brain process relations.

1.5.2. Mind-Full Study

The log data collected from the Mind-Full Nepal study consists of different forms of data: signal quality from the EEG headset that was used, the Neurosky relaxation (R) and attention (A) index values (0–100), Mind-Full game performance data (e.g., game ID, game start time, game finish time, number of tokens collected, threshold value for R/A, hold time value) and headset calculated EEG data (in brainwave bands).

The over-arching research questions for this thesis are:

(RQ1) What are the appropriate measures and statistical methods to analyze the log data from the Mind-Full Nepal study to understand patterns in the data across the sessions? [Methodological Contribution]

(RQ2) Did the children's ability to self-regulate their relaxation and attention change (improve) across their Mind-Full training sessions? [Applied Contribution]

RQ1 sets the stage for the contribution of this thesis. RQ2 is further divided into specific research questions related to different approaches in measuring relaxation, attention, EEG patterns and game performance within groups across the sessions. These are further discussed in Chapter 5.

1.6. Structure of the Thesis

In Chapter 2, I discuss the fundamentals of how EEG works, and describe the relationships between different brainwave activity and mental state of the brain. Then, I outline how to estimate the relaxation (affective brain state) and attention (cognitive brain state) from the brainwave amplitude based on previous research. Then, I summarize studies that investigated the effectiveness of the Neurosky headset. Following this, I survey the NF research that had a study design like the Mind-Full Nepal study. This survey provides a summary of relaxation and attention measures, and the statistical methods used to analyze these measures. In Chapter 3, I present my side study, completed on a Mitacs internship. It was a controlled experiment conducted with adults to validate the effectiveness of the NMM consumer-grade EEG headset in detecting relaxed/anxious state and attentive/non-attentive state. In Chapter 4, I briefly set the context for the data from the Mind-Full Nepal study by explaining the Mind-Full system, study design, and details about the data collected. This will help the reader gain more clarity about the contribution of the current thesis (in the context of the larger Mind-Full project), which is primarily the development and application of the data analysis approach for dealing with change across sessions. In Chapter 5, I present my detailed research questions, define the measures I have developed, and outline my approach to data analysis. I also discuss the rationale behind each approach, their constraints and limitations. In Chapter 6, I present the results for each research questions. In Chapter 7, I discuss the appropriateness of my analytical approach, based on my interpretation of the results and previous research works. I also discuss the limitations of my research, and suggest alternate methods, recommendations, and scope for future work. In the Conclusion, I summarize the contributions and primary outcome of this thesis.

Chapter 2. Background

2.1. Overview: Brain Anatomy and Functionality

In this section, I give a brief description of the functionalities of the brain relevant to this thesis. The brain acts as an agent and director of different physical, mental, and emotional activities of the body. The brain is the source of “emotion, cognition, memory and intelligence.” (Baars & Gage, 2010).

2.1.1. Regions of the Brain

The brain consists of three different parts – the cerebrum, cerebellum and brain stem.

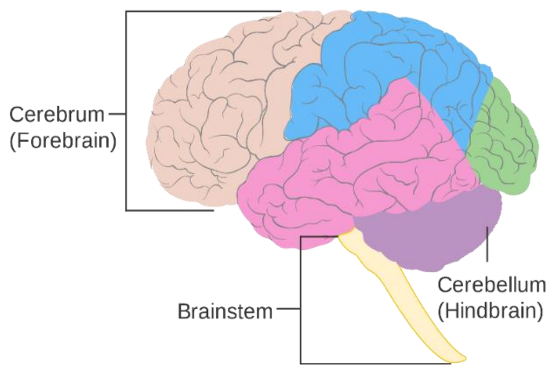


Figure 2.1. Different parts of the brain.[2]

Cerebrum or Cerebral Cortex

The cerebrum[3] or the cortex is the largest part of the brain. It is responsible for a person’s thoughts and actions. It has four lobes: frontal lobe, parietal lobe, occipital lobe and temporal lobe. The frontal lobe is partly responsible for regulating emotions and reactions, problem-solving, attention, concentration, reasoning, and some elements of movement. The parietal lobe manages movement, orientation, perception, and

[2] https://commons.wikimedia.org/wiki/File:Diagram_showing_some_of_the_main_areas_of_the_brain_CRUK_188.svg
[3] http://serendip.brynmawr.edu/bb/kinser/Structure1.html

recognition. The occipital lobe is responsible for visual perception. The temporal lobe is responsible for auditory stimuli, memory, and speech. The temporal lobe contains the amygdala, which is the seat of fight or flight responses.

Cerebellum

The cerebellum is located below the cerebral cortex, and it takes care of motor movements, muscle movements, and balance. (Teplan, 2002)

Brain-Stem

The brain-stem consists of the pons, medulla, and mid-brain. It acts as a connector between the cerebral cortex and cerebellum, and connects these parts with the spinal cord[4]. The brain-stem is responsible for autonomic events such as respiration and digestion (Teplan, 2002).

2.1.2. Why is pre-frontal cortex important to the current research?

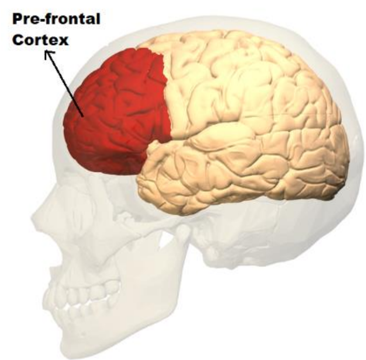


Figure 2.2. Illustration of prefrontal cortex in the brain[5].

The frontal lobe consists of the pre-frontal cortex (PFC), pre-motor cortex, and motor cortex. The pre-frontal cortex is responsible for the executive functions (EF) of a human being, which include the regulation of affective and cognitive processes like anxiety and attention (Goldstein, Naglieri, Princiotta, & Otero, 2014).

[4]http://www.mayfieldclinic.com/PE-AnatBrain.htm
[5] Modified from the source: https://commons.wikimedia.org/wiki/File:Prefrontal_cortex_(left)_-_lateral_view.png

Executive Functioning

Executive functioning in childhood includes the principal elements of "anticipation, goal selection, planning, initiation of activity, self-regulation of emotion, mental flexibility, deployment of attention, and utilization of feedback" (Anderson, 2002). Best et al. (2009) define executive functioning as "an umbrella term to encompass goal-oriented control functions of the PFC." The executive functions involve "self-regulation of emotions" and "deployment of attention." Self-regulating emotions include self-monitoring to find out if a person is overwhelmed by minor things, anxious, over-stimulated or unable to calm down, which can potentially hinder the goal completion. Deployment of attention includes preventing issues such as "running out of steam before homework is done, switching between tasks without completion, daydreaming and not being focused in present activity"[6]. It is important to monitor the PFC activity while children are trying to self-regulate anxiety, relax/calm down, and be attentive to a task. In this research, I explore patterns of relaxation or anxiety and attention or inattention using EEG data collected from children's pre-frontal cortices.

How can we monitor the PFC?

Activity in the PFC, like any other region of the brain, can be measured through techniques such as EEG sensors and FMRI (functional magnetic resonance imaging)[7]. FMRI is not suitable as a measurement instrument for field studies as it involves large equipment which cannot be deployed easily at schools. EEG[8] is a non-invasive technique that can easily be deployed beyond lab settings to sense the activity of the PFC region of the brain. EEG sensors record electrical signals that are released through the skull from different regions of the brain. The electrical signals are generated when groups of neurons fire in synchronous activity, which tend to correspond to specific brain processes. These electrical signals can be classified into different brainwaves based on their frequencies. This is discussed further in Section 2.2.

[6] http://mentalhealthdaily.com/2014/04/10/beta-brain-waves-12-hz-to-40-hz/
[7] http://fmri.ucsd.edu/Research/whatisfmri.html
[8] https://en.wikipedia.org/wiki/Electroencephalography

2.2. Relaxation and Attention Measurements from EEG Data

One can learn if a person is relaxed/anxious or attentive/non-attentive by sensing and interpreting the neurological electrical signals generated in the cerebral cortex of the person's brain. In this section, I discuss the generation of the electrical signals from the nerve cells in the cerebral cortex region (in Section 2.2.1), the different frequencies of these electrical signals, what affective or cognitive states they correspond to (in Section 2.2.2), and how relaxation (affective state) and attention (cognitive state) can be measured from these different frequencies of the electrical signals (brain waves) (Section 2.2.3).

2.2.1. Neuron Activation and Brainwaves

The neurons or nerve cells in the brain communicate with each other through electrical activity. When a neuron is electrically charged, it transmits positive and negative ions through its cell walls. Billions of neurons are involved in exchanging charged ions from one neuron to another throughout the cerebral cortex – forming a wave. The ions are transmitted to the electrodes of an EEG sensor when the sensor is placed on the head. This transmission of ions among the neurons escapes the scalp and can be captured as a charge in electric potential (in volts). This electrical activity can be sensed and recorded over time through a non-invasive method by the EEG medical imaging process[9]. The EEG is applied to the cerebral cortex region. Electrical signals generated by large numbers of nerve cells are generally weak when they reach the outer layer of the scalp. However, they are amplified and then stored digitally on microchips embedded in the headset and/or sent directly to computers for analysis. (Teplan, 2002).

2.2.2. Brainwaves

Brain-based neural electrical signals (or brainwaves) have two components: frequency and amplitude. Brainwaves (collections of synchronous neural activity) occur at frequencies that typically range from 2 to 50 Hz (Teplan, 2002). Amplitude refers to the strength of the signal for each brainwave frequency. Amplitude is measured in microvolts.

[9] https://en.wikipedia.org/w/index.php?title=Electroencephalography&oldid=761415384

Both frequency and amplitude of brainwaves reflect different brain states of a person. The brainwave power, generally proportional to the square of the amplitude, is also analyzed in order to understand different brain states. The scope of this literature review is to understand the change in relaxation and attention with respect to change in amplitude of different brainwaves. This research uses the collected brainwave amplitude calculated by the Neurosky proprietary algorithm. This is further discussed in Section 2.2.3.

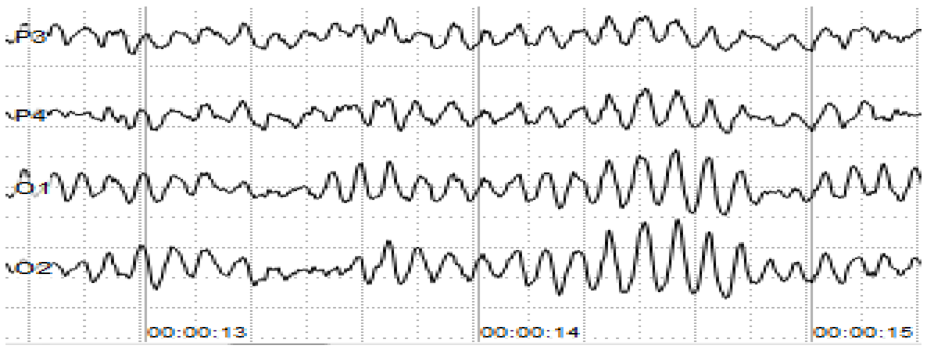


Figure 2.3. EEG sample recorded at different scalp locations. Horizontal axis depicts time, and vertical depicts amplitude (uV). [10]

Alpha Waves (8 - 13 Hz)

Alpha waves range between 8 and 13 Hz and occur during EC relaxation under awake conditions (Antonenko, Paas, Grabner, & van Gog, 2010). Alpha waves diminish when a person is intensively attentive (e.g., arithmetic calculation) and strengthen as soon as the subject stops concentrating. This is known as 'alpha blocking' (Empson, 1986). As alpha waves are closely associated with relaxation, an increase in the amplitude of alpha waves is considered to be the desirable outcome in biofeedback training (Empson, 1986). Even though alpha waves are pre-dominant during EC conditions, alpha waves can be improved by increasing relaxation through the NF technique under an EO condition (Dempster & Vernon, 2009; Putman, 2000). Alpha amplitude is more in the posterior region compared to the frontal region, but alpha amplitude was found to be higher in all brain regions under non-directive meditation conditions (Lagopoulos et al., 2009). Alpha-enhancing biofeedback training was widely used in relaxing and improving the conditions of anxiety patients (Moore, 2000).

[10] https://commons.wikimedia.org/wiki/File:Human_EEG_artefacts.png

Beta Waves (12 - 30 Hz)

Beta wave frequencies[11] range between 12.5 and 30 Hz and they are classified as: low beta waves or Sensorimotor Rhythm (SMR) (12.5–15 Hz), mid-range beta waves (15–20 Hz), and high beta waves (18–40 Hz). Low beta waves occur during a relaxed yet attentive state. Mid-range beta (15–20 Hz) waves correspond to focused performance and mental calculations, and high beta waves correspond to high energy or arousal, stress, tension, and anxious states. The increase of low beta waves in PFC was observed during relaxed yet attentive states (Belkofer & Konopka, 2008). For people with ADHD, NF treatment involves improvement of low beta or SMR (12–15 Hz) (Gruzelier, Egner, & Vernon, 2006; Ogrim, Kropotov, & Hestad, 2012). Some studies have considered mid-range beta for attention enhancement (Bakhshayesh et al., 2011; Lubar et al., 1995).

Theta Waves (4–8 Hz)

Theta waves[12] oscillate between 4–8 Hz. Theta waves are mostly linked to drowsiness, meditation (Shin, Lee, Shin, & Shin, 2014), and hypnotism, and occurs during REM sleep (Moorcroft, 2013). Theta activity is observed in meditation, especially for experienced meditators and/or deep meditation (Broughton & Hasan, 1995; Lagopoulos et al., 2009). An increase in ADHD symptoms was associated with increase in theta amplitude and NF treatments aim to reduce theta amplitudes (Bakhshayesh et al., 2011 ; Gruzelier et al., 2006 ; Lubar et al., 1995). This brainwave is dominant in children, people with ADHD, and individuals who are unable to focus on a task. There are further details on these studies in the forthcoming sections.

Gamma Waves (30 - 50 Hz)

Gamma brainwaves are the fastest brainwaves (30–50 Hz). These waves are generated during expanded consciousness, self-realization, and spiritual emergence. They are related to the presence of altruism and higher virtues[13].

[11] http://mentalhealthdaily.com/2014/04/10/beta-brain-waves-12-hz-to-40-hz/
[12] http://mentalhealthdaily.com/2014/04/12/theta-brain-waves-4-hz-to-8-hz/
[13] http://www.brainworksneurotherapy.com/what-are-brainwaves

Delta Waves (1 - 4 Hz)

Delta waves occur during dreamless sleep and deep meditation where a person lose external awareness[14]. Healing and regeneration of the body and mind occur during this process, which explains the necessity of deep sleep. They are considered the slowest brainwaves (1 to 4 Hz) and are responsible for the involuntary processes of the body such as heart regulation, and digestive and renal functioning[15].

Table 2.1. Association between occurrence of a brainwave and their corresponding mental/psychological state.

Brainwaves	Related State of Mind
Alpha waves (8–13 Hz)	Relaxation, meditation
Low beta waves or SMR (12.5–15 Hz)	Attention, concentration
Mid beta waves (13–20 Hz)	Attention, complex mental task, focused performance
High beta waves (20–30 Hz)	Anxiety, stress, high energy
Theta waves (4 - 8 Hz)	Daydreaming, drowsiness, attention deterioration
Gamma waves (30 - 50 Hz)	Self-realization, spiritual emergence
Delta waves (1 - 4 Hz)	Deep sleep (when external awareness is lost) and involuntary processes, such as heart regulation.

2.2.3. Brainwaves and their Corresponding Affective and Cognitive States

In this section, I present the implications of NF feedback/physiological studies conducted with both adults and children to provide evidence that EEG brainwave amplitudes can be used to detect anxious, relaxed, and attentive states.

Relaxation

Alpha brainwave biofeedback was used as a therapy for adults with anxiety and depression disorders by medical professionals (Stinson & Arthur, 2013). Alpha brainwaves are predominant in mindfulness practice where the goal is to attain relaxation and to

[14] http://www.brainworksneurotherapy.com/what-are-brainwaves
[15] http://mentalhealthdaily.com/2014/04/14/delta-brain-waves-0-hz-to-4-hz/

improve “creativity, decision making and problem solving skills” (Stinson & Arthur, 2013). Similarly, both alpha and theta waves were found to be higher in experienced meditators, while performing non-directive meditation technique compared to eyes-closed (EC) relaxation (Lagopoulos et al., 2009). Egner et al. (2002) studied the NF alpha/theta training with university students to induce relaxation. In this research, the author suggested that the alpha activity improves with the onset of relaxation and when the participant moved from relaxed state to drowsy state, the theta amplitudes becomes more dominant than alpha amplitudes. A study conducted using a audio-visual biofeedback technique to improve the relaxation of 18 athletes found that the mean alpha (8–13 Hz) amplitude improved for the training group in comparison to the control group (Mikicin & Kowalczyk, 2015). Similarly, an anxious state can also be figured out with an increased theta/alpha ratio or decreased alpha activity. NF techniques were deployed to train alpha activity (8-13Hz) for adolescents and children who had depression, anxiety disorders, and PTSD (Simkin, Thatcher, & Lubar, 2014). Even though some of the previous studies had been conducted under the EC condition, few of those studies investigated the alpha amplitude under the eyes opened (EO) condition. In adults, the alpha activity was found to decrease under task performance compared to resting in both EC and EO conditions (Legewie, Simonova, & Creutzfeldt, 1969). An EO biofeedback relaxation training conducted with 77 army veterans showed that there was an increase in the alpha amplitude across sessions, and decrease in the beta and theta activity (Putman, 2000). Similarly, EO alpha-based biofeedback training with 29 adults across 10 sessions showed that the mean alpha amplitude significantly improved across sessions (Dempster & Vernon, 2009). From these studies, I infer that I can expect an increase in the alpha amplitude (8-13 Hz) during the EO biofeedback training for relaxation.

Attention

In most of the biofeedback interventions developed for ADHD, treatments aim to decrease the theta (4–8 Hz) amplitude and beta amplitude (12–20 Hz) (Bakhshayesh et al., 2011 ; Leins et al., 2007; Lubar et al., 1995). Belkofer et al. (2008) showed that there can be a decrease in the delta and theta amplitude and an increase in beta amplitude in the left PFC of an adult, implying an enhanced focus while performing a creative task. For children, the characteristic protocol for ADHD aims at increasing the amplitude of SMR (12–15 Hz) (Kropotov et al., 2005; Russell-Chapin et al., 2013,). An increase in cognitive focus can also be determined by a decrease in alpha activity (Antonenko et al., 2010). Leins et al. (2007) and Gruzelier et al. (2006) suggested that the measures in NF training for ADHD children include a decrease in the amplitude of theta (4–8 Hz), an increase in the amplitude of beta (12–15 Hz), and an increase in the amplitude of SMR (12–15 Hz). Ogrim et al. (2012) studied the increase in the theta (4–8 Hz) amplitude, decrease in the beta (12–20 Hz) amplitude, and increase in theta/beta ratio to detect ADHD in children and adolescents. Chabot et al. (1996) suggested that the theta (4–8 Hz) amplitude is elevated in children with ADD in comparison with typical children in the frontal region. For ADHD, NF training improves the amplitude of SMR (12–15 Hz) and beta (15–18 Hz) (Fuchs et al., 2003). Thus, for children, I can expect a possible increase in the low beta or SMR (12–15 Hz) and/or the beta range (12–20 Hz), and a decrease in the theta amplitude (4–8 Hz) for improvement in attention.

Summary

As seen in the previous section, several studies have children as participants where the aim of the study was to improve the attention of the children. The dependent measures of these studies match with the dependent measures of the attention enhancement studies conducted with the adults. For relaxation, there is little research with children as participants when conducting the NF training to decrease anxiety or improve the relaxation and most of the relaxation-based studies were conducted with adults. Nevertheless, I can consider the dependent measures used in these studies and adapt them to my current thesis. According to Simkin et al. (2014), the core brain development occurs in the early stages of life and continues to remain same through the adolescence and adults.

For this thesis, I decided the dependent variables such that they are consistent with the above-mentioned NF studies. However, it is important to note that some of these frequency bands overlap with each other. For example, alpha, a measure of relaxation is from 8-13Hz whereas the SMR, the measure of attention is from 12-15Hz. With that, the range of frequencies given by the NMM headset slightly differs from the standard range used in the above studies. Similarly, there might be a personality or individual differences on how mental states change with brainwave patterns, depending on age and individual factors (Gruzelier, 2014). So, for the measures of relaxation and attention, I used frequency ranges that are consistent with the previous research with the children. These studies show that the amplitude of alpha can increase with relaxation. The amplitudes of alpha (8–13 Hz) and theta (4–8 Hz) can increase with meditation techniques, especially for experienced meditators. The amplitudes of beta (12–20 Hz) and SMR (12–15 Hz) increase during focused attention, while the amplitude of theta goes down with an increase in attention. The current thesis deals with the data collected from non-experienced meditators (i.e., children). Therefore, an increase in the alpha (8–13 Hz) amplitude for EO relaxation is expected. For the EO attention task, I can anticipate an increase in the amplitude of SMR (12–15 Hz) and low beta (15–18 Hz), and a decrease in the amplitude of theta (4–8 Hz). Since the alpha range is from 8 Hz to 13 Hz for Neurosky EEG amplitude values, I can consider the increase in the amplitudes of their low alpha (7.5–9.25 Hz) and high alpha (10–11.75 Hz) as a measure of the increase in relaxation. For an increase in attention, an increase in the amplitudes of their low beta (13–16.75 Hz) and a decrease in their theta (3.5–6.75 Hz) are expected.

Table 2.2. Neurosky headset measures for relaxation and attention.

Dependent measures from previous research	Corresponding NMM measure	Mental states
Increase in amplitude of alpha (8–13 Hz)	Increase in amplitude of low alpha (7.5–9.25 Hz) and high alpha (10–11.75 Hz)	Improvement in relaxation
Increase in amplitude of SMR (12–15 Hz) and beta (12–20 Hz)	Increase in amplitude of low beta (13–16.75 Hz)	Improvement in attention
Decrease in amplitude of theta (4–8 Hz)	Decrease in amplitude of theta (3.5–6.75 Hz)	Improvement in attention

Note. Table deals with change in amplitude and corresponding change in mental/psychological state.

2.3. Review of Consumer-grade EEG Headsets

In this section, I present the advantages of using consumer-grade headsets. Then, I introduce the consumer-grade headset NMM, which was used for the Mind-Full study. Then, I discuss those studies that have validated the proprietary algorithm of the NMM headset.

2.3.1. Advantages of Consumer-grade EEG Devices Compared to Medical-grade EEG Devices

Affective and cognitive states such as relaxation, calmness, and attention can be captured from the frontal region of the brain (Liu, Chiang, & Hsu, 2013). Brainwaves can be collected in a controlled laboratory environment using medical-grade devices, such as the international 10–20 electrode placement system[16]. In this system, 37 saline-dipped electrodes are placed on the scalp to record brainwave amplitudes. The electrodes can capture an accurate recording of the scalp signals but is inconvenient for usage outside the laboratory environment. However, this set-up reduces users' movement and comfort, and restricts the place in which the recording can be done (Johnstone, Blackman, & Bruggemann, 2012). Such set-ups would be difficult to use, especially on children with PTSD and ADHD, as they require the participants to have restricted movement. The inconveniences of using the medical-grade, multi-sensor EEG recording methods and the advantage of determining affective states in the frontal region led to the invention of the less-invasive consumer-grade EEG sensors. Though these devices are less accurate than the 10–20 electrode system (Searle & Kirkup, 2000), they are more convenient to use for day-to-day purposes. Therefore, there is a trade-off between accuracy and convenience. These consumer-grade EEG devices are simple, small, portable, and wireless, and thus easier to use compared to the traditional EEG devices. Unlike the medical-grade EEG devices, the consumer-grade EEG devices use dry electrodes for real-time transmission of the brainwaves to a smartphone or a computer.

Consumer-grade EEG devices have their proprietary algorithms for translating brainwave patterns into affective and cognitive scores such as attention, meditation, frustration, and engagement, based on the brainwave frequency and amplitude. These scores are potentially used as controllers in human-robot interaction (Vourvopoulos et al., 2012; Vourvopoulos et al., 2014), for NF in everyday gaming systems (Antle et al., 2015) and as attention regulators in learning systems (Huang et al., 2014). Use of consumer-grade EEG devices makes such systems readily available, easy to use, and inexpensive. In the consumer market, there are several consumer-grade EEG headsets such as

[16] https://en.wikipedia.org/wiki/10-20_system_(EEG)

Muse[17], Emotiv Insight[18], and Neurosky MindWave Mobile (NMM)[19]. They are readily available in online markets and stores, and are affordable (price range is from $99 to $300) compared to research-grade EEG systems. They also provide software support in different operating systems, which encourages the development of new NF applications.

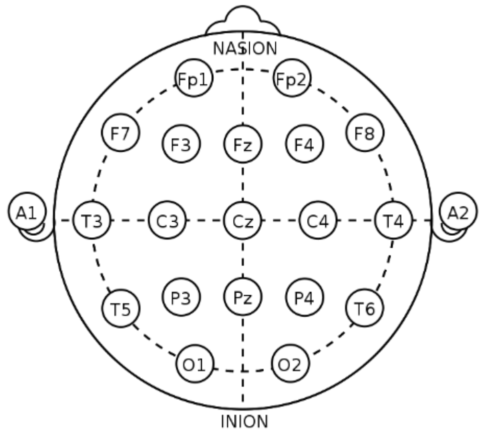


Figure 2.4. Electrode locations in 10–20 electrode system. [20]

2.3.2. Neurosky MindWave Mobile (NMM)

The Neurosky MindWave Mobile (NMM) uses a single dry electrode (Fp1) on the left prefrontal lobe to determine attention and meditation. The meditation score represents “the level of mental calmness or relaxation” and is the inverse of anxiety or stress. The attention score represents the “user's level of mental focus or attention, such as that which occurs during intense concentration and directed (but stable) mental activity” (Masasomeha, 2017). NMM is a comparatively affordable headband ($99 CAD), which can be readily purchased from online stores, and it provides software support to develop new software applications in both mobile (Android and iOS) and PC platforms. NMM can be connected to a smartphone, tablet or Personal Computer (PC) via Bluetooth. Once connected, the NMM proprietary algorithm called eSense Meter transfers information such as timestamp, signal quality, relaxation score, attention score, eye blink strength, and EEG power spectrum representing the relative strength of each brainwave frequency (Masasomeha, 2017). This information is transferred at the rate of 1 sample per second

[17] http://www.choosemuse.com
[18] https://www.emotiv.com/insight
[19] http://store.neurosky.com/pages/mindwave
[20] https://en.wikipedia.org/wiki/10-20_system_(EEG)

to the connected device along with raw EEG at the rate of 512 samples per second to the connected PC or smartphone. This raw data can be used to produce desired brainwave frequency using Fast Fourier Transform (FFT) method (Sałabun, 2014). I provide further details in Chapter 4 about the data collected for the Mind-Full Nepal study from the NMM. For the Mind-Full game, the NMM was chosen. This is because the NMM satisfied the requirements of the Mind-Full game: it records relaxation, attention, and relative EEG amplitudes; it can be programmed to work with the Android tablet environment; it was the only commercial headset approved for use with children at the time of study; and finally, it is robust and non-invasive (easy to put on) for children, portable, wireless and affordable – which makes it suitable for use in a field environment like Nepal (Antle et al., 2015).

Figure 2.5. Neurosky mindwave mobile headset[21]. The electrode is positioned in the left side of the forehead.

2.3.3. Previous Work that Validated Neurosky's Data Quality

Several previous studies were conducted to test the validity of Neurosky's proprietary algorithm. In a controlled experiment with 34 participants, the attention score from the NMM headset positively correlated with the self-reported attentiveness (Mendez

[21] Modified from the source: https://www.flickr.com/photos/digitalgamemuseum/7973437584

et al., 2009). However, they did not study the relaxation element. The NMM headset was found to be effective for detecting the overall change in relaxation and attention by matching the self-reported state of mind (calm or stressed) level with individual percentage time spent below relaxation scores 40, which indicates stress (Crowley, Sliney, Pitt, & Murphy, 2010). However, the methodology was unclear as Crowley et al. (2010) did not explain why they did not analyze the attention scores despite collecting them, and why they did not run the inferential statistics on relaxation and attention scores. Researchers did not find any difference in NMM's relaxation and attention scores between visual-intense and relaxation tasks (Stinson & Arthur, 2013). Here, the EC/EO condition and data analysis procedure were not clearly discussed. Rogers et al. (2016) used the NMM headset to detect the EEG conditions during three minutes of EC, three minutes of EO, and three minutes of a visual cognitive task counterbalanced with 19 youths, 21 adults, and 19 elderly participants. The results revealed that there was a significant increase in alpha activity and a decrease in theta activity in the EC condition. Similarly, there was a decrease in alpha activity and an increase in beta activity when they compared EO and visual cognition conditions. The study did not validate the relaxation and attention scores from the headset, as their aim was to understand if the EEG patterns of Neurosky were consistent with EEG patterns described in the previous studies for EC and EO conditions. Johnstone et al. (2012) conducted a comparative validation study between (1) the power spectrum of the raw data collected from the research grade 10–20 EEG system and (2) the raw data collected from the NMM's headset from 20 adults with no mental health issues. The study revealed that there was a strong positive correlation between the two power spectra of the two headsets. Even though the authors had discussed the use of 'relaxation' and 'attention' score given by the headset, they did not perform a comparative validation of these scores attained during each task. Robbins et al. (2014) conducted a validation study to investigate the efficacy of EEG amplitudes, relaxation and attention indices provided by the Neurosky's proprietary algorithm. This study was conducted with 24 participants with tasks that clinically induced relaxation, attention and required working memory. This study showed that the relaxation index was significantly lower while performing the task than when compared to a resting state. For attention, instead of an objective measure, they correlated the attention score with self-reported attention level. All the other EEG amplitudes related to relaxation and attention were not significant between the tasks. There is a gap in the research. Commercial-grade EEG headset

NMM's effectiveness in measuring affective and cognitive brain states has not yet been established because the previous works did not statistically determine if the algorithm can differentiate between relaxed and attentive states. In this thesis, I address this research gap in Chapter 3 as a tertiary contribution.

2.4. Review of Neurofeedback Data Analysis Approaches

To analyze the log data from the Mind-Full Nepal study (Antle et al., 2015), I need to understand how different analysis methods were used in previous research on NF log data. In this section, I explore the data handling techniques used in previous NF research in order to develop an approach to the analysis of patterns in NF data across sessions. For each paper, the research goal, study design, dependent variable(s), constraints such as handling noisy and missing data (if described), analysis method, results, and limitations are summarized. Finally, I highlight the data analysis method chosen for my study and discuss how the past studies helped me choose this method.

In the previous studies, the trend of relaxation and attention across the sessions were analyzed in two different ways. Some studies have considered (almost) all the sessions continually (e.g., from session 1 to session 30 in a 30-sessions study or from session 1 to session 16 in a 20-sessions study) for analysis. Other studies have considered discrete sessions for inter-sessional analysis (e.g., Session 1,4,9,13 or session 1 and 20). Some have considered blocks of discrete sessions for analysis (e.g., comparing the mean DV of sessions 2,3,4 with session 19,20,21). In Section 2.4.1, I discuss the studies which have considered discrete sessions (or) blocks of discrete sessions for across-session analysis. In section 2.4.2, I discuss the studies which have considered continuous sessions for across-session analysis. In Section 2.4.3, I describe a study that considered all the sessions as different blocks for analysis. In these sections, apart from the session sampling strategies, I also discuss pre-processing methods, such as handling noise and missing data. In Section 2.4.4, I describe NF works that did not have a study design similar to the Mind-Full study but dealt with other related methods such as data pre-processing techniques and dependent variables derived from consumer-grade EEG headsets. Finally, in Section 2.4.5, I summarize important aspects of these studies that are applicable to the Mind-Full data analysis approach in this thesis. The review of this section is summarized in *Table 2.3*; this table summarizes the studies discussed below, focuses on how they sampled the sessions and handled data pre-processing, and presents the statistical method used in each study. Green represents studies where continuous sessions were sampled. Blue represents studies that sampled discrete sessions.

Table 2.3. Table of dependent variables, no. of sessions sampled, data pre-processing, and statistical analysis method used in previous NF studies.

Author, Year; No. of Participants	Dependent Measure	Session sampled for analysis	Data pre-processing (if mentioned)	Data Analysis With IV
(Lee, 2009) *N* = 14	Mean relaxation; mean attention; N/A – S	N/A	Data with 50% or more noise were excluded	One-way ANOVA with groups (novice/ experts) as IV
(Lim et al., 2012) *N* = 20	Mean attention score per session;	24 sessions in total. Session 1 and 20 sampled for analysis.	Data with noise and children who dropped out (or) having missing sessions were removed.	Paired *t*-test with session (1 and 20)
(Thomas et al., 2013) *N* = 5	Mean attention score, mean accuracy;	3 sessions in total; Session 1 and 3 sampled for analysis.	No information	Descriptive statistics and graphs for 1st and 3rd sessions. Paired *t*-test for accuracy for session 1–3.
(Stinson & Arthur, 2013) *N* = 13	Mean relaxation score, mean attention score from headset; Mean of Neurosky calculated EEG amplitude	N/A	No information	Independent *t*-test with type of training (visual relax and auditory relax)

Author, Year; No. of Participants	**Dependent Measure**	**Session sampled for analysis**	**Data pre-processing (if mentioned)**	**Data Analysis With IV**
(Lubar et al., 1995) *N* = 19	Mean theta amplitude;	30 sessions. All considered for analysis	Participants were classified based on their performance.	Spearman correlation for mean theta amplitude vs 30 session (for individual participants)
(McDonald, 1974); *N* = 10	Mean alpha frequency	10 sessions; All considered for analysis	No information	Mixed ANOVA with training/control group and sessions (10) as IV
(Kirenskaya, Novototsky-Vlasov, & Zvonikov, 2011)	Log transformed brainwave amplitudes (to meet normality)	N/A	For normality, log transformation was done	Mixed ANOVA, Greenhouse Geisser correction with electrode level (10), groups (2) as IV
(Egner et al., 2002) *N* = 18	Mean theta/alpha;	5 sessions; All considered for analysis	Taken care at study design level.	Mixed ANOVA with groups and sessions (5) as IVs
(Dempster & Vernon, 2009) *N*=29	Mean alpha amplitude. Percentage time above threshold	10 sessions; All considered for analysis.	Four participants who had missing sessions were removed	Repeated measures ANOVA
(Gevensleben et al., 2014) *N* = 10	Mean amplitude of SCP (+ and -).	13 sessions in Total; Sessions 1,5,9,13 sampled for analysis	Noise and eye blinks were handled at software level.	Mixed ANOVA with sessions (1,5,9, 13) and polarity (positive, negative) As IV.

Author, Year; No. of Participants	**Dependent Measure**	**Session sampled for analysis**	**Data pre-processing (if mentioned)**	**Data Analysis With IV**
(Drechsler et al., 2007) *N* = 30	Mean amplitude of SCP, Mean amplitude of negative SCP;	15 double sessions. Double sessions (2–3) and double sessions (14–15) sampled for analysis	Missing data was substituted with group mean.	MANOVA with positive and negative SCP as DV; double sessions (2–3) and double sessions (14–15) as IV;
(Bakhshayesh et al., 2011) *N* = 30	Mean amplitude of theta/beta;	30 sessions; 3 blocks of sessions sampled for analysis. (1 block = 10 session).	Data with noise were excluded. Children who dropped out at the end were still included for last block.	Mixed ANOVA with 3 blocks of sessions game type as IV
(Hillard et al., 2013); *N* = 10	Mean amplitude of theta, beta and alpha;	12 sessions	Data with noise and eye blinks were excluded	Repeated measures ANOVA with sessions (12) as IV
(deBeus & Kaiser, 2011); *N* = 42	Engagement Index (EI), pre-post behavioural scores;	40 sessions; First three and last three sessions sampled for analysis.	Participants were separated as good performers and poor performers for analysis.	Pearson correlation between change in EI (last 3 sessions – first 3 sessions) and change in pre-post behavioural scores (only for good performers)
(Takahashi et al., 2014); *N* = 9	Mean amplitude change of SCP;	20 sessions; 16 sessions sampled for analysis.	To include all participants, last few sessions were excluded.	Repeated measures ANOVA with sessions (16).

Note. N/A means across session analysis was not performed. These studies dealt with within-session analysis.

2.4.1. Discrete Session Sampling for Between-Session Analysis

In this section, I discuss the NF studies that have considered discrete samples of sessions for analyzing the trend in inter-session analyses. Each study is described in the following order: goal and study design, task, DV, session sampling, data pre-processing, data analysis method, result, limitation (and/or my critique), and implications for Mind-Full data analysis.

Drechsler et al. (2007) conducted 15 sessions of NF training with 17 children in a NF training group and 13 children in a control group. The control group was given standard behavioural therapy. The EEG was recorded at the scalp position Cz at the rate of 250 Hz. The cortical potential can be regulated to reduce ADHD symptoms. The intervention is designed to move from negative to positive slow cortical potential (SCP) using visual stimuli of colours, with red indicating increase in cortical excitability and blue indicating decrease in cortical excitability. The task consists of two seconds of baseline reading (i.e., data collected when no task was performed), six seconds of feedback (using cartoons), and six seconds of transfer (where no feedback is given). For analysis, sessions (2,3) and sessions (13,14) were considered as Drechsler et al. (2007) wanted to investigate the improvement in the DVs at the end of the training compared to the beginning of the training. For missing data, the group mean score was substituted by the researchers. Repeated Measures (RM) MANOVA were conducted with mean positive SCP and negative SCP as DVs and session (2,3) and session (13,14) as IV. The results revealed that the mean amplitude of negative SCP showed significant change in ending sessions (13–14) compared to beginning sessions (2–3) in both trainings. Even though this study is not relevant to our current thesis, it is important to understand how they sample their sessions. Instead of considering all the sessions, they grouped the first two and last two sessions' means to look for improvement in the treatment over time. In terms of data pre-processing technique, the method of substituting mean value for missing data is notable, however, substituting group mean might not be relevant as each participant could be different.

Lim et al. (2012) designed a 24-session study (eight weeks; three sessions per week; 30 minutes per session) to evaluate attention-focused NF training for 20 un-medicated ADHD children (aged 6–12, *M* = 7.8, SD = 1.4). The task was to control a 3D avatar using the attention score from the single channel dry EEG sensor, NMM. The attention score was proportional to the speed of the running avatar. The parents filled an ADHD rating score (for inattentiveness and hyperactive symptoms) before treatment at the end of the 1st, 4th, 8th, 20th and 24th sessions. Three follow-up sessions were conducted after 24 sessions (one session per month). Session 1 and 20 were considered for analysis to understand the improvement in mean attention score. The influence of noise was removed from analysis by filtering data points wherein the corresponding NMM's noise parameter indicated the presence of noise. The mean attention scores were calculated after this noise exclusion. Six participants were removed as they either dropped out or had missing sessions. Paired *t*-test with mean attention score per session as DV and session (1 and 20) as IV revealed that the mean attention score did not significantly vary from the 1st session and 20th session. Spearman correlation revealed that there was a negative correlation between change in attention score and change in ADHD (difference between sessions 1 and 20). I do not agree with their session sampling technique. Comparing the change in DV for two sessions when they had approximately 20 sessions may not be representative of all the NF sessions. In this paper, it is important to note that they excluded data points with poor signal quality (given by the NMM headset). Likewise, they excluded children who dropped out of the study from the analysis (data pre-processing).

Thomas et al. (2013) studied the effectiveness of a three-session NF game to enhance attention and memory among five children using Emotiv EPOC headsets. The attention level of the participants was collected during the resting phase to determine the baseline. During the NF training, a matrix of numbers would appear for two seconds and subsequently, certain numbers from the matrix would disappear. The task was to remember and recollect the missing numbers. Only when the attention level was above the baseline would the game allow the user to fill the missing numbers. To determine the effectiveness of the NF training, accuracy, time taken, and attention score (calculated from the strength of raw signal through their algorithm) were gauged. Even though they had three sessions, only the 1st and 3rd sessions were considered for analysis. No details on data pre-processing were given. A paired *t*-test with percentage accuracy as DV, and sessions (1 and 3) as IV showed that there was a significant improvement in accuracy in last session in comparison with the first. The descriptive statistics and individual plots showed gradual improvement across all three sessions, but they did not run inferential statistics with data from all three sessions. The first and last session were sampled for their analysis (Session sampling).

deBeus et al. (2011) conducted NF training using a video game with 42 children (aged 7–11) with ADHD. The study design had three assessments (pre-mid-post) and two training sets of 20 sessions each (each session lasting 30 minutes) before and after the mid-assessment. After the pre-assessment, the children were randomly split into two groups that interchanged two conditions (NF training and placebo-control) for the first and second set of trainings. The signal is recorded at Fz position using 10–20 systems. The engagement index (EI; calculated as amplitude of beta/ (theta + alpha)) was used to control the car speed in the NF game. The first three and last three sessions were considered for analysis. In the training group, the children were divided into NF learners (31 out of 42) and NF non-learners (11 out of 42) for data analysis purpose. Participants were considered as NF-learners if the difference in mean EI of the last three sessions was 1.5 SD greater than the baseline sessions (first three sessions). For NF-learners, Pearson correlation between change in behavioural score and change in EI showed significant correlation for the training group and no correlation for the placebo group. This study supports two ideas: Session sampling, the difference in mean of the first few and last few sessions can be correlated with pre-post behavioural scores and data pre-processing, grouping participants based on their performance for analysis. However, it would have been more a comprehensive analysis if they had reported the change in EI from all the participants before concentrating on only those who performed well.

Gevensleben et al. (2014) discussed their methodological framework in using NF as a treatment for children with ADHD by conducting a study with 10 boys (age 10–13) in 13 sessions (one to three sessions per week; each session was 105 mins). In the NF training, children were asked to regulate their SCP in two ways – negative SCP (for attention purpose) and positive SCP (for relaxation). EEG activity was recorded, at position Cz, for the 1st, 5th, 9th, and 13th sessions with an interval of four sessions between session samples. Reward points were given on successful completion of NF training; however, it was not clearly mentioned how often it was given or the nature of the visual cues used for NF. Behavioural assessment by parents (using standard ADHD questionnaires) were done before and after each training. The paper does not discuss missing data, noise or outliers. A mixed ANOVA was conducted with sessions (1, 5, 9, 13), polarity (positive, negative) as IV, and mean SCP as DV. The result implied that the children reached negative mean amplitude during negative training, but failed to reach positive mean amplitudes during positive training. This work shows another way of sampling sessions; however, they did not provide justification for their sampling method.

2.4.2. Continuous Sessions Sampling

In this section, I describe those studies that have sampled continuous sessions for the analysis.

Lubar et al. (1995) studied if NF was effective in reducing ADHD symptoms over 30 sessions (five sessions per week, 8 to 10 weeks) with 19 children with ADHD (mean age 11.8 years). They used 10–20 electrode systems to collect EEG data from positions Pz, Cz, and Fz with Cz reference at the rate of 128 Hz. Each training session included two minutes of baseline period (to calculate the threshold amplitude), five minutes of NF reading, and five minutes of NF listening. Theta amplitude and beta amplitude were collected. Thirty sessions were considered for analysis. No information on missing or noisy data handling was provided. Spearman rank order correlation for mean theta amplitude vs sessions revealed that there was a negative correlation (implying improvement in attention) for 12 subjects (later grouped as EEG change group) and no correlation for seven subjects (no EEG change group). Reporting that 12 children had a reduction in theta amplitude and seven children did not have a reduction in theta amplitude does not provide evidence if the NF significantly helped the children in reducing ADHD symptoms. This study provides us with two data analysis strategies: sampling continuous sessions for analysis and the data pre-processing method of grouping children as responders and non-responders.

McDonald (1974) conducted a randomized controlled trial to study if the alpha-based NF training was effective in improving task focused skills in children (N = 30; aged 9–10 years). From previous research, the author investigated if the alpha frequency NF training improved alertness in daily activities and prolonged interest and focus on task. He further investigated if this alertness helps children, who generally struggle to hold prolonged interest and focus on tasks. In his study, children were classified into three groups: the first group had an auditory NF based on their alpha brainwave (8–13 Hz), the second group had soothing music in auditory feedback for the same alpha brainwave, and the third group received no treatment. The alpha frequency was recorded with a research-grade EEG headset. Pre- and post-test on auditory and visual stimuli memory tests were conducted, and time-on-task and accuracy were recorded. Ten sessions were conducted, with each session lasting for 10 minutes. All 10 sessions were included in the analysis. It is difficult to assess the reliability of the results as the authors did not provide details on how they handled outliers, noise, and missing data, especially when 10 continuous sessions were considered for the analysis. Mixed ANOVA was performed with group and sessions as IV, and mean percent time in alpha band and mean dominant alpha frequency as DV. Based on insignificant results, the author concluded that the alpha NF training alone could not improve focus and it needed to be supported by other classroom interventions. Even though analyzing frequency is out of scope for the current thesis as I only have EEG amplitudes, the important takeaway is consideration of NF data from continuous sessions to investigate change over time.

Egner et al. (2002) studied whether alpha/theta NF training was effective for helping alcoholics. Eighteen participants (8f; 12m; age M = 23.1, SD = 1.9) with no prior experience in feedback training were divided into training (9 participants; 7m; 2f) and control (9 participants, 5m; 4f) groups. A research-grade EEG device was used to collect raw signals from a Pz electrode at a sampling rate of 256 Hz, which was fed to the band-pass filter to extract alpha (8–11 Hz) and theta (5–8 Hz) bands. The study comprised of five sessions, 15 minutes each, at the rate of two to three sessions per week. All five sessions were considered for data analysis. It is difficult to assess the reliability of the results as the authors did not provide details on how they handled outliers, noise, or missing data. A 2 × 5 factor mixed ANOVA was performed with group and session number as IV, and mean T/A ratio per session as DV. There was no significant change in mean T/A ratio across sessions. The mean T/A ratio per session was higher in the training group compared to the control group. This is another study where all the continuous sessions were considered to look for a change in the pattern of mean T/A ratio.

Hillard et al.(2013) studied the effect of NF on 10 children (age *M* = 13.6, *SD* = 3.5) with ADHD for 12 sessions (25 minutes each) to improve attention. The NF was in the form of visual and auditory information feedback, which was triggered when the focus level dropped below the baseline. The focus level was given by a third-party software. The raw EEG signal was collected from prefrontal region FPz at 128 Hz and filtered. Relative amplitude per minute was calculated for each of the brainwave frequencies, theta (4-8 Hz), low beta (13-18 Hz), and alpha (8-13 Hz) with respect to total signal (from 2 to 50 Hz) present at that minute. All 12 sessions were considered for the analysis. Those data points with excessive eyeblinks were removed from the analysis. No information was provided regarding excluding outlier, missing sessions or noise. Repeated measures ANOVA were conducted with all 12 sessions as IV, and mean relative amplitude of theta, low beta, and alpha per session as DVs. The results showed that the reduction of mean amplitude of theta, theta/beta, theta/alpha, and improvement for mean amplitude of low beta over sessions was significant, implying improvement in attention. The author indicated that they included all 12 sessions for the analysis but did not clearly explain their data handling methods such as outlier and noise. This lack of clarity may cause reliability issues and makes it difficult to reproduce the study methods. From this work, I can determine that the significant improvement in attention was observed by considering 12 continuous sessions for analysis.

Takahashi et al.(2014) conducted an SCP training with nine children with ADHD for 20 sessions of 12 minutes each. The task was to excite the cortical variability to positive and negative condition by moving the visual stimuli up/down. The SCP was recorded at Cz electrode at 128 Hz. For the analysis, they considered 16 continuous sessions (out of the 20 for analysis). This is because one of the children who participated in the study had only 16 sessions. Another child who quit halfway through the study was not included in the data analysis. (data pre-processing). In this way, the authors could include most of the data collected for the analysis. Separate repeated measures ANOVA were conducted for both positive and negative conditions with SCP as DV and session number (*N* = 16) as IV. For both negative and positive conditions, the main effect of the sessions was significant on the mean SCP. Like the previous study, this study considered continuous sessions instead of only considering the first few and last few sessions to identify the trend of the dependent measure across sessions, and showed significant results. This study also informs my analysis about determining an inclusion and exclusion criteria for participants based on the number of available sessions for each participant.

Dempster et al. (2009) conducted an alpha NF training on 29 adults for 10 sessions (one per week) under EO condition to study the change in alpha activity (8-12 Hz) at Pz for both within and across sessions. In each session, the baseline alpha was collected for three minutes under EO rest conditions. After this, two seven-minute NF training was conducted where the participants had to improve their alpha amplitude above the baseline and the feedback was given in the form of a visual cue. The DVs collected were the alpha amplitude, percentage time above the baseline and the integrated alpha measures (percentage time x amplitude/100). Ten sessions were considered for the analysis. Some participants dropped out of the study after five sessions and they were excluded from the analysis (data pre-processing). For across sessions analysis, repeated measures ANOVA was conducted for each of the DVs and sessions as IV. There was a significant main effect of the session on the alpha amplitudes, however the percentage time and the integrated alpha were not significant. Even though the author claimed that the alpha amplitude improved over the sessions, when I take a closer look at their analysis, only the ninth session had significantly higher amplitude than the first session and the rest of the sessions had non-significant gradual improvement. This study informs us that the alpha NF can be potentially achieved under EO conditions through visual feedbacks, which is similar to the Mind-Full relaxation tasks.

2.4.3. Grouping of Sessions as Blocks

Bakhshayesh et al.(2011) conducted an NF training with 35 children (6–14 years each) for 30 sessions (2–3 sessions per week, 30 minutes each) to improve attention. The children played three different games where they collected game points. The treatment group (NF) received points based on theta and beta amplitudes and the control group was rewarded points based on facial muscular movements (EMG). For the analysis, 30 sessions were aggregated into 3 blocks of sessions (10 sessions per block). They explained that they followed this method of sampling to comprehend the change in dependent measure. However, no proper explanation was given on how grouping sessions as blocks would provide better observation of the change in the dependent measures. Data with head movements, headset disconnection and technical errors were excluded. Three children who did not cooperate well had their data excluded. One child who dropped out after the 25th session was included in the analysis (data pre-processing). No details were provided on any outliers or missing data substitutions. A mixed ANOVA was conducted with three blocks of sessions as IV1 and games as IV2 for each of the groups separately. The results revealed that the training group had significantly decreased theta/beta (*T*/*B*) ratio across the three blocks. The control group showed significant reduction in EMG, below baseline, across blocks. This study presents us with another method of analysis where the sessions can be grouped into blocks, and the blocks can be analyzed with repeated measures. Categorizing sessions into blocks might reduce the fluctuating variability in the data. However, by using this method of session sampling for my RQ, it may be difficult to observe session to session patterns in the data (e.g., gradual improvement).

2.4.4. Other Methodologically Relevant NF Studies

In this section, I present other relevant NF studies that may inform us about handling data pre-processing strategies. These studies do not have a similar study design to my Mind-Full study design. However, these studies have analyzed NF data similar to that of the Mind-Full study.

Kirenskaya et al.(2011) studied the waking EEG powers of 30 healthy participants (age 19–52; *M* = 34.4, SD = 10) using 10–20 EEG systems under the EO condition. Recorded EEG raw data were filtered into different brainwaves. Before performing statistical analysis, the brainwave powers were subjected to natural logarithm to make the distribution close to normality. A repeated measures ANOVA was performed with groups (high-hypnotized, low-hypnotized), electrode positions, hemisphere (left, right) as IVs and the natural logarithmic transformation of EEG powers for each brainwave as DV. The study revealed significant difference between the EEG powers between high and low hypnotized subjects and waking restful conditions for theta and alpha bands. In this NF study, it is important to note that the author performed a log transformation of the DVs when the data was not normally distributed (data pre-processing). In this way, they were able to reduce the skewness in the data and make the data fit into normal distribution.

Stinson et al. (2013) conducted a study with 15 participants practising audio and video feedback using NMM to see if there was a significant difference in relaxation score, attention score, and EEG amplitudes given by the Neurosky. A paired *t*-test was conducted for each DVs (mean relaxation score, mean attention score, individual EEG mean amplitudes) with the type of feedback training as IV. Results revealed that there was no significant difference in mean relaxation score, mean attention score or mean alpha amplitude between the audio and visual training approaches. Beta amplitudes (both low and high) were significantly higher for the video tutorial compared to alpha state training. However, the study did not provide details such as how they handled noise or if the experiment was conducted with EO/EC conditions. The paper claimed that the non-significance in the data was because the mean relaxation and mean attention score were mostly around 50, and it was homogenous for most of the participants. This was the only study to use Neurosky calculated EEG amplitudes instead of raw signals. The study results are important for my current thesis as they have the same DVs as mine.

Lee (2009) studied the attention and relaxation levels of 14 archers (9m; 5f, age: 11 to 40 years, experience: 0.5 to 15 years) using NMM headsets. The archers were asked to shoot their targets at short (18 m) and far distances (70 m). They shot around 15 to 20 arrows. Based on the headset's noise parameter, shots with 50% or more noise were excluded. Mean attention and relaxation scores were calculated at each shot (time period) and grand mean was calculated for every archer. Improvement over each shot was observed through mean attention and mean relaxation score per shot and their slope with time using linear regression. However, no information on the statistical significance or correlation coefficients was provided, which makes it difficult to assess the claim of improvement over sessions. This study informs my analysis that there can be an exclusion criterion based on noisy signals. However, Lee could have just excluded the data with noise as done in other studies instead of excluding the entire session (data pre-processing). The measures used in this study are like the measures used in the Mind-Full study. Thus, the method for calculating mean relaxation and attention score per session is important to consider (DV selection).

2.4.5. Summary

In this section, I summarize different approaches to the data analysis such as dependent measures (also discussed in Section 2.2.4), data pre-processing methods (e.g., noise/outlier exclusion, exclusion criteria for missing data, data transformation, if necessary), session sampling, and inferential statistics for inter-session analysis. I also summarize the pre-processing methods discussed in my literature review. The details on how these approaches were adopted to my Mind-Full data are discussed in Chapter 5.

Dependent Variables

From the above studies, for across-session analysis, we see that the dependent variables were either mean amplitude of different brainwave frequencies or mean NMM index scores (such as attention, relaxation, etc.) calculated by the headset for each session. A significant improvement of mean relaxation and attention across sessions would suggest that the children have improved their ability to relax and attend (as represented by their relaxation/attention scores) over the course of the Mind-Full sessions. This would support the gains reported from the behavioural data analysis (Antle et al., 2015), and could increase the claims that Mind-Full has aided in self-regulation. As stated earlier, however, a lack of improvement could be due to several external factors and improvement across sessions is not the only common pattern. When relaxation and attention scores were available from the EEG headsets, the mean relaxation per session or the mean attention per session were considered as DVs (e.g., Lee, 2009; Lim et al., 2012; Stinson & Arthur, 2013; Thomas et al., 2013). In other studies that have collected EEG amplitude information in different bands, the mean amplitude per band per session or a block of sessions were used as dependent measures depending on the affective or cognitive state being studied (e.g., Bakhshayesh et al., 2011 ; Egner et al., 2002 ; Hillard et al., 2013 ; Lagopoulos et al., 2009 ; J. F. Lubar et al., 1995). Therefore, for my analysis, for every session, I will calculate mean relaxation, mean attention, and mean of Neurosky calculated EEG amplitudes (low alpha and high alpha as relaxation metric; theta and low beta as attention metric) as DVs.

Data Pre-Processing and Constraints

For data pre-processing and constraint handling techniques (such as handling missing or noisy data), most of the works did not mention if there were any missing data or outliers, how it was handled except a few studies where it was mentioned it briefly. With respect to noise exclusion, studies that have used a research-grade EEG headset removed noisy data (Bakhshayesh et al., 2011; Hillard et al., 2013; Lim et al., 2012) Most of the studies that have used the NMM headset had not discussed noise exclusion (Crowley, Sliney, Pitt, & Murphy, 2010; Robbins & Stonehill, 2014; Rogers et al., 2016). Lee (2009) excluded the whole session if there was high noise level (indicated by the NMM) above 50%. For missing data, Drechsler et al. (2007) substituted group mean for missing values, which can be problematic depending on how many values are missing and the variation of the data. Takahashi et al. (2014) skipped participants if they had completed less than half of the sessions compared to others. They selected enough sessions so that most of the participants were considered while including most of the sessions for analysis, unlike other studies where only the first and last session(s) were included. For violation of normality, Kirenskaya et al. (2011) used log transformation of NF data when assumptions of normality were not met. The log transformation can be done for continuous ratio data when normality was not met because the data was highly skewed (Keene, 1995; Lane, 2017). In general, transformations reduce the variability in the data and might not represent patterns in the original untransformed data (Feng et al., 2014). Therefore, for this thesis, in case of non-normal distributions that did not fit a standard transformation method (e.g. not skewed), I decided to run the analysis with untransformed data to reduce the chance of not seeing patterns or trends across the sessions. In Chapter 5, the procedures used to handle these data constraints are discussed. For the data pre-processing, I derive my rationale either from these previous works (if they have discussed) or I reason out my own rationale based on common statistical procedures.

Sampling Sessions

For repeated measurement analysis, the sessions were dealt with in three different ways. Firstly, in some studies, a selected single session was considered for analysis (Gevensleben et al., 2014; Lim et al., 2012; Thomas et al., 2013). However, we need to consider the variations in children’s performance due to individual, family or community level events (e.g., trauma) over the course of their sessions. In addition, the NMM can provide noisy data and the quality of data may vary based on the sensors (e.g., I discovered that some children's dirty foreheads were leaving dirt on the sensor and quality was deteriorating over sessions). Selecting a few sessions, with gaps between sessions, might be an issue because the selected session might be an outlier or non-representative. This phenomenon is discussed by Zuberer, Brandeis, and Drechsler (2015).

Secondly, some studies have grouped the sessions as blocks that were used for analysis (Bakhshayesh et al., 2011; Drechsler et al., 2007). This is a useful way of studying the data, where we group sessions into blocks for analysis, thus smoothing inter-session variability. Drechsler et al. (2007) grouped the first few and last few sessions as two blocks for analysis and they excluded all the middle sessions from analysis. This type of analysis would address the question of whether the children improved their self-regulating abilities at the end of the training when compared to the beginning. However, this approach would not detect the pattern of improvement, which is of interest to me.

Thirdly, some studies have considered almost all the continuous sessions in their analysis (Egner et al., 2002 ; Hillard et al., 2013 ; McDonald, 1974 ; Takahashi et al., 2014). Some of these studies showed significant improvement across sessions. As a first step in my analysis, it would be wise to include continuous sessions (except outliers) instead of sampling only the initial and final sessions to observe if there is an improvement in relaxation and attention with time as the children practice the game.

Statistical Analysis

I need to choose a statistical analysis method that could identify significant variability in the DV that is caused by two factors: sessions (repeated measurements) and groups (between-subject factor). In my literature survey, four different statistical methods were used to analyze changes in patterns of the DVs for repeated measurements. They are repeated measures ANOVA (Hillard et al., 2013; Takahashi et al., 2014), mixed ANOVA (Bakhshayesh et al., 2011; Egner et al., 2002; Gevensleben et al., 2014; Kirenskaya et al., 2011; McDonald, 1974), regression (Lee, 2009), and correlational analysis of session and DV for individual participants (Lubar et al., 1995). Apart from sessions, I also have a between-subject factor, 'groups,' for the Mind-Full study. In this case, both mixed ANOVA and regression method can address the variability caused by both session and groups. The regression model assumes that the DV and the predictors has a relationship (e.g., linear) and there is a possibility that these patterns might not be present in all my datasets. In addition, performing linear regression with categorical variables as predictors would provide similar results that we can get from repeated measures ANOVA[22]. Therefore, based on the previous study, I chose to perform 2 × 22 mixed ANOVA to identify significant patterns of the dependent measures across the training period with groups (G1 and G2), and sessions (S2 to S23) as independent measures.

[22] http://www.theanalysisfactor.com/why-anova-and-linear-regression-are-the-same-analysis/

Chapter 3. Headset Validation Study

3.1. Introduction

In this section, I describe an independent study that I did to investigate the effectiveness of the consumer-grade EEG headset NMM in detecting relaxed state (from anxious state) and attentive state (from relaxed state) during a short duration of time (3 minutes per task). This study was not a part of the Mind-Full study and was conducted independently for my Mitacs internship with Wearable Therapeutics Inc., Vancouver, BC. The study was important for the company, as they wanted to use the NMM to find the amount of time spent in relaxed and attentive states by using the self-regulating intervention manufactured by them. However, the results of this study may inform the validity of using the NMM headset in the Mind-Full Nepal study.

Consumer-grade EEG devices have their own proprietary algorithms to calculate affective and cognitive scores (e.g., attention, meditation, frustration, engagement, and memory) from the brainwave frequency and amplitude. These scores can potentially be used as controllers in human-robot interaction (Vourvopoulos et al., 2012; Vourvopoulos et al., 2014), for NF in everyday gaming systems (Antle et al., 2015), and as attention regulators in learning systems (Huang et al., 2014). Use of consumer-grade EEG devices make these systems readily available, easy to use, and inexpensive (Huang et al., 2014). Therefore, for the success of systems, it is important to know if the proprietary algorithms of consumer-grade EEG devices can reliably measure affective and cognitive states. Although previous research has dealt with validating the reliability of the consumer-grade EEG devices, there is a gap in the literature in objectively validating the proprietary algorithms of consumer-grade EEG devices in detecting the affective and cognitive states. In this study, I investigated the effectiveness of the proprietary algorithm of the consumer-grade EEG headset - Neurosky MindWave Mobile (NMM) in detecting relaxation and attention of the user. The criteria for choosing the headsets were: ease of use, portability, compatibility with different operating systems, affordability, extensibility (i.e., the software could be used as a basis for new systems), availability in consumer market, and use of dry electrodes.

The NMM uses a single dry electrode to determine attention and meditation. The meditation score represents "the level of mental calmness or relaxation" and as an inverse of stress, and the attention score represents "the user's level of mental 'focus' or 'attention,' such as one which occurs during intense concentration and focused (but stable) mental activity" (Masasomeha, 2017). Muse uses dry electrodes to determine mellowness and concentration. Mellowness is a measure of relaxing state, and concentration is defined as "focusing on something particular, thinking about something with intensity. . .trying to solve a problem, or working your intellectual mind"[23]. The Emotiv EPOC[24] is a 14-channel research-grade headset. EPOC determines a person's affective states, such as relaxation and frustration, and cognitive states such as boredom and attention. Even though studies had shown that EPOC is fairly reliable with 14 EEG sensors (Vourvopoulos et al., 2014), it uses saline-dipped electrodes instead of dry electrodes. Emotiv Insight[25] is a 5-channel dry electrode headset that helps to read the brainwaves from five sensors, and determines the user's levels of meditation, stress, attention, and enjoyment.

As of August 2015, the Emotiv Insight software support was still under development, and the headband was not readily available in the consumer market. Both Muse and NM are readily available in online markets and stores, comparatively affordable, and provide software support to various operation systems which encourages developing NF training intervention. Thus, I initially selected these two headsets for my study. I categorized mellowness (from Muse) and meditation (from NMM) as measures of users' relaxation, and categorized concentration (of Muse) and attention (of NMM) as measures of users' attention. Although I conducted the study with the Muse headset in 2015 and collected the data, the Muse headset manufacturer, Interaxon Inc., instructed us not to use their relaxation and attention indices for analysis as it was only experimental data, and the indices were removed from their later versions of the software (Moffat, email communication, April 2017). The headset validation study has the research question: "Is the proprietary algorithm for NMM headset effective in detecting users' relaxation (or stressed state) and user's attentive state (or non-attentive state)?" I used three tasks that

[23] http://developer.choosemuse.com/research-tools/available-data
[24] https://www.emotiv.com/epoc/
[25] https://www.emotiv.com/insight/

can induce calmness, attention or stress. The IV was the type of task: eyes closed (EC) relaxation, a math test, and Stroop test. The DV was the percentage of time when the algorithm detected relaxation and attention.

Based on the evidence from the previous research (explained later in section 3.3.4), I assumed that EC relaxation might make the participant more relaxed and less anxious. The Stroop test has been shown to clinically induce anxiety and attention. The math test can clinically induce anxiety and attention. I predicted that relaxation would be higher in EC relaxation, and would subsequently drop in math or Stroop tests, showing an increase in stress. Conversely, I predicted that attentiveness would be low in EC relaxation when compared to math and Stroop tests. The results of this study will be useful for researchers who use the off-the-shelf Brain Computer Interface (BCI) for their systems and consumers of commercial BCIs who wish to self-regulate their affective and cognitive states.

3.2. Existing Research on Validating Consumer-Grade EEG Devices

In a study with 34 participants, the attention score from the NMM headset were positively correlated with self-reported attentiveness (RebolledoMendez et al., 2009). Crowley et al. (2010) found that the NMM headset was effective for detecting the overall change in meditation and attention by matching the self-reported affective state with individual relaxation scores. In their study, the methodology was unclear on how they collected data for the relaxation task, and no inferential statistics were performed on the relaxation and attention measures. In a study conducted by Stinson et al. (2013), no significant difference was found in the relaxation and attention score of NMM between visual-intense and relaxation tasks. However, they did not clearly explain the EC/EO condition and data analysis procedures. There is a gap in the research. No researcher has objectively validated the affective and cognitive scores calculated by commercial-grade EEG headsets. It is important to conduct this work to ensure that future BCI and NF work is based on accurate detection algorithms. In my study, I adopted an objective methodology to validate the relaxation and attention scores calculated by the proprietary algorithms of the consumer-grade EEG headset – NMM.

3.3. Methodology

3.3.1. Study Design

The research design was a *within-subject counter balanced study* with headset type as independent variable. Nine participants were first exposed to NMM and then to the Muse headset while the remaining nine participants were first exposed to Muse, followed by the NMM headset. It is important to note here that the Muse headset was included in the study design, method, and data collection procedure, but during the data analysis stage, we decided to exclude the Muse data.

3.3.2. Participants

The participants (N=18, M=8, F=10) were adults (average age =27 years, SD = 3.5). The participants themselves declared that they did not have any medical history of developmental disorders or anxiety disorders, and they were tested for non-colour vision deficiency. The participants did not have any prior experience or practice in any relaxation or mindfulness techniques. The ethics review for this study is in Appendix B.

3.3.3. Research Instruments

The Muse has four electrodes (Tp9, Fp1, Fp2 and Tp10) and three reference electrodes. The headband gives a mellow (i.e., relaxation) and concentration value (both ranging from 0.0 to 1.0) at the rate of 10 Hz[26]. The NMM uses a single dry electrode (Fp1), and a reference electrode. NMM gives an attention and meditation value (both ranging from 0 to 100) every second through their proprietary algorithm eSense (Masasomeha, 2017). According to Neurosky, a meditation score of 40 or below indicates anxiousness or stress, and an attention score of 40 or below indicates non-attentiveness (Masasomeha, 2017).

[26] http://developer.choosemuse.com/research-tools/available-data

3.3.4. Tasks

To evaluate the headsets, I selected three tasks: EC relaxation, timed math test, and timed Stroop test. According to Neurosky documentation, EC relaxation in an idle case (i.e. not performing any complex mental activity) will promote relaxation (Masasomeha, 2017). Liew et al. (2014) and Meece et al. (1990) suggested that an arithmetic task can induce anxiety. Stroop test can be used to clinically induce anxiety (Leite, Maria De Lourdes, Sartori, & Andreatini, 1999; Prinsloo, Derman, Lambert, & Rauch, 2013) and attention (Robbins et al, 2014; Stinson et al, 2013). Based on this research, I predicted that the EC relaxation task would promote a relaxed state for the participants and that the math and the Stroop tests would reduce relaxation and increase the attention of the participants, when compared to EC relaxation.

In EC relaxation, participants were asked to breathe and relax with their eyes closed. Melodious instrumental music was played in the background. The math test was a timed test where the participants were asked to solve 100 simple arithmetic questions (e.g., 5*8? 49/7?) in three minutes. Consequently, it demanded the participants' concentration to complete the task. In the Stroop test, the participants had to identify the colour of the text, where the text itself was spelled in a different colour. For this study, I used a computerized Stroop test which lasted three minutes. In both the tests, the participant's score was presented to them in order to motivate their constant focus on the test. The time was displayed using a digital stop clock.

3.3.5. Procedure

The participants were asked to remain seated in a closed room, wearing one of the headbands. To calibrate the algorithm and establish acceptable headset connectivity, no data was recorded for the first two minutes. In the first task, the participants followed EC relaxation for three minutes. Subsequently, the participants took a timed computerized math test, where they were asked to solve 100 math questions in three minutes. Finally, the participants took a timed computerized Stroop test for three minutes. The headsets were cleaned with rubbing alcohol before the start of the experiment. The participants were instructed not to talk or make any sudden head movements during the task. After a 10-minute break, the tasks were repeated with the other headset, in the same order. I did not counter-balance the order of the task as I felt that it would be difficult for participants to switch from an anxious to relaxed state within the three-minute time-frame, especially when they did not have any previous experience with relaxation/mindfulness techniques, and this could be a possible confounding factor on the relaxation task scores.

Assumption: Stroop and math test makes a person more anxious/attentive compared to EC relaxation.

(RQ0.1) Does the EEG headset NMM show more time spent in relaxation while performing a clinically relaxing task compared to an anxiety-inducing task?

(RQ0.2) Does the EEG headset NMM show more time spent in attention while performing a cognitively challenging task compared to a relaxing task***?***

Why: To understand if the NMM can clearly differentiate the time spent on the attentive and non-attentive states, and the time spent on the relaxed and non-relaxed states under clinical conditions.

3.3.6. Data Collection

I used an Android app to collect data (at one sample/second) from NMM using Neurosky's software developmental kit (SDK). The SDK provides the data that includes the relaxation score, attention score, and signal quality. I used Interaxon's research tools, Muse-IO, and Muse Lab to collect data from the Muse headset. The Muse-IO establishes connection between the headset and the app. and the Muse Lab stores the relaxation value, attention value, brainwave potentials, signal quality, and headband contact status. The headsets were newly bought and tested before conducting the study.

3.3.7. Data Analysis

For noise exclusion criteria, noisy data comprised data points where the corresponding Poor_Signal was greater than zero. By removing noise in the data, I was able to filter out those relaxation and attention indices that were a duplicate value from a previous time instance or had zero as its value. After noise exclusion, the percentage time spent above the relaxation and attention level of 40 was calculated for every participant, similar to the approach considered by Crowley et al. (2010). The percentage time above the relaxation or attention index 40 = [(Total number of seconds where relaxation or attention index is greater than 40 when Poor_Signal = 0) / (Total number of seconds when Poor_Signal = 0)] * 100. Here, the parameters, Poor_Signal, and the relaxation and attention indices were sampled at 1 Hz. For example, if total time spent on the EC relaxation task was 180 seconds, and the person spent 170 seconds above the relaxation threshold of 40, the percentage of time was calculated as [(170/ 180) *100]. Therefore, the DV is the Percentage of Time spent above the 40% threshold, for both relaxation and attention, and the IV is the task (relax, math, Stroop). Since this was a within-subject study, three measurements of the same type were recorded for three different tasks. After checking the assumptions, repeated measures ANOVA or Friedman tests were used to analyze the change in percentage time spent above threshold 40 for the tasks on each headset. Outliers were excluded using a box-plot (1.5 times IQR) and were substituted with the mean percentage time of that task. Note: The results from the Muse headset is not reported here, as their relaxation/attention indices were experimental and reported as obsolete in their latest version of their software (Interaxon company, personal communication, April 2017).

3.3.8. Results

Relaxation

Three outlying data-points (Figure 3.1.) were excluded and substituted with the mean percentage time above relaxation score 40 of the tasks of all participants.

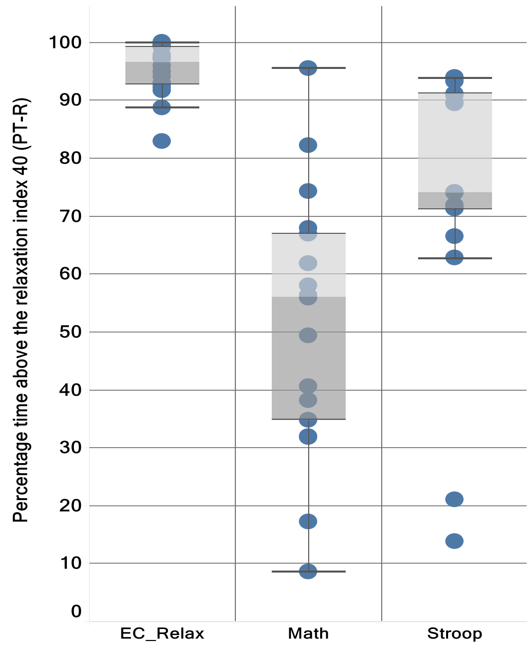


Figure 3.1.   Percentage time above the relaxation index 40 (PT-R) vs. type of task (*N*=18). Box plot was used to detect the outliers.

The Shapiro Wilk test revealed that the normality was met for the percentage time for EC relaxation and math test ($p > 0.05$), and normality was not met for the Stroop test ($p < 0.05$). The Mauchly's test revealed that the sphericity of variance was violated, $p < 0.05$. Therefore, the Friedman test was conducted to analyze the change in percentage time spent above relaxation score 40 (PT-R) with the type of the task. The result revealed that there was a significant difference in the PT-R, which significantly changed for the type of the task (N=18, $\chi^2(2)$ =28.44, $p < 0.001$, *Kendall's W* = 0.79). For the post-hoc, Wilcoxon signed rank test was conducted. The result revealed that the EC relaxation had significantly higher PT-R than the math test (Z= -3.72, $p < 0.001$, r= 0.62) or the Stroop test (Z= -3.41, p < 0.001, r= 0.56). The Stroop test had significantly higher PT-R than the math test (Z = - 3.54, $p < 0.001$, r= 0.59). Therefore, as expected, the NMM showed higher PT-R spent on relaxation for the relaxing task when compared to the cognitively challenging tasks. Even though I did not have any expectation about the difference in PT-R between the math and Stroop tests, it is interesting to note that the Stroop test had significantly higher PT-R than the math test.

Table 3.1. Descriptive statistics for relaxation measure (*N*=18).

Type of task	Median (IQR)
EC relaxation	*96.51 (92.91 to 99.29)*
Math test	*56.07 (34.07 to 67.2)*
Stroop test	*81.16 (71.27 to 91.67)*

Attention

Two data points were removed as outliers and substituted with mean PT-A of all the participants for that task.

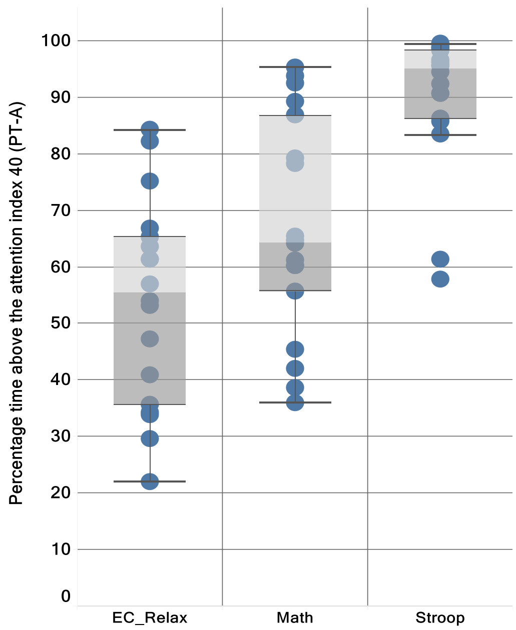


Figure 3.2. Percentage time spent above the attention index 40 (PT-A) vs. type of task (*N*=18). Box plot was used to detect the outliers.

The assumptions of normal distribution and sphericity of variance were met, *all* $p > 0.05$. Repeated measures ANOVA was conducted with type of task as IV and PT-A spent above attention threshold 40 as DV. The result revealed that the type of task had a significant effect on the percentage of time spent above attention threshold 40 with high effect size, N=18, $F\ (2,34) = 29.52$, $p < 0.001$, $\eta^2 = 0.63$. The post-hoc test (Bonferroni) revealed that the PT-A was significantly higher for the Stroop test when compared to EC relaxation with high effect size, ($p < 0.01$, $r = 0.81$) and math test with moderate effect size ($p < 0.01$, $r = 0.64$). The math test did not have a significantly higher PT-A when compared to EC relaxation ($p > 0.05$, $r = 0.21$). Unlike what I expected, when compared to EC relaxation, only the Stroop test had significantly higher PT-A, and even though the math test trended toward higher PT-A, the difference was not statistically significant. Therefore, I cannot definitively claim that the NMM has the potential to differentiate attentive state from the relaxed state.

Table 3.2. Descriptive Statistics for attention measure (*N*=18).

Type of task	Descriptive
EC relaxation	*M* =53.72, *SD* =18.20
Math test	*M* = 67.1., *SD* = 19.49
Stroop test	*M* = 93.81, *SD* = 4.83

3.3.9. Discussion

In this study, 18 participants performed EC relaxation, a math test and a Stroop test for three minutes each. From the results, I can infer that the NMM can distinguish relaxation (from an anxious state) in terms of time spent on relaxation for the task. For attention, we could not make a strong claim as the result was significant between EC relaxation and Stroop test whereas it was not significant between EC relaxation and the math test. This might be due to the math anxiety faced by the participants, which might have inhibited their attentive state. According to Neurosky documentation, increase in anxiety might inhibit attention (Masasomeha, 2017).However, the descriptive statistics showed a trend in which the math test had a higher percentage time than EC relaxation. Due to this, no strong claim could be made for attention. Based on my results, I infer that the NMM may be a potential research instrument in distinguishing relaxed state from anxious state. Our result agrees with the validation study conducted by Robbins et al. (2014) who observed higher relaxation score and no significant difference for the attention index for the rest state when compared to the problem solving state This result contradicts the study conducted by Stinson et al. (2013), who did not find any significant difference in relaxation and attention. However, they considered the relaxation and attention indices as continuous data instead of considering them in terms of percentage time.

The result of this study statistically verified the claims of the previous research discussed earlier, where the effectiveness of the NMM algorithm was validated by analyzing self-reports with percentage time spent below threshold 40 (Crowley et al., 2010; Rebolledo-Mendez et al., 2009). From my results, I can claim that the NMM can be a potential research instrument in detecting relaxation for controlled experiments conducted with adults as participants. However, it is important to understand that the study does not validate or claim the accuracy of the relaxation or attention scores (e.g., scores were provided based on EEG or facial muscular movements), and that results may not be applicable to children.

Limitations

I originally included the Muse headset in the study because the Muse headset was one of the prominent ones available in the consumer market. Even though I collected the data from the Muse headband, I do not report the results here because the Muse relaxation/attention indices, which were under development when I began the study, were later reported as obsolete. If this information had been available before I began, I would have limited the study to the NMM headset alone by excluding the Muse headset from the study design.

Another limitation is that the intensity of the affective states (relaxed, stressed) experienced by our users may have been reduced when they performed the same tasks with the second headset. I compensated for this issue by counterbalancing the order of the headsets. However, with a small sample size, this might not be adequate. The tasks did not provide any feedback on the relaxation or attention score to the participant during the study, therefore the result might not hold well for NF studies.

In this study, I considered the percentage of each task time spent over the R/A thresholds of 40 as the DV, even though the data given by the headset was continuous. Although mean relaxation/attention have been widely used in previous studies (see section 2.4), I decided to use the percentage time spent above the threshold as used by Crowley et al. (2010). This informs how long the participant stayed in relaxed and attentive states in each task rather than an average value that may have large variations. However, the results could vary if the continuous data was analyzed instead of the percentage time spent above a threshold.

The results might not hold for people from different age groups (e.g., children and elderly). Thus, this result cannot be directly applied to the Mind-Full Nepal study, where participants were children and the study was conducted in a field environment. In addition, since we controlled factors such as high noise, user movement, and actions while conducting the study, the study's results might not generalize to a real-world or non-laboratory setting (e.g., a school in Nepal) where it may be more difficult to control the user's movement and actions while wearing the headsets. This would require follow-up with a field study.

Contribution and Future Works

The result of this study demonstrates that in a controlled environment, the NMM headset is capable of distinguishing mental states (relaxation and attention) that are relevant to the Mind-Full game. The study provides a small methodological contribution that can be used in future studies to evaluate commercially available headsets' algorithms for detecting affective and cognitive states in lab conditions. In this study, I did not collect raw EEG signals. For future work, it would be intriguing to correlate the relaxation and attention measures from the headsets with the actual brainwave relaxation (alpha amplitude) and attention (low beta, theta amplitude) measures to better understand if the NMM is accurate in detecting relaxation and attention.

3.3.10. Conclusion

In this study, I statistically validated if the proprietary algorithm of NMM detects and distinguishes relaxation and focus in 18 adults. My results showed that NMM is mostly effective in distinguishing relaxed/anxious. For attention, we got mixed results and could not make a stronger claim. The results of this study are important as it is important to assess the validity of the proprietary algorithm of commercially available headsets used in BCIs to ensure accuracy of the functionality of consumer-grade NF systems.

Chapter 4. Mind-Full Study Scenario

In this chapter, I will briefly describe the research instrument, study design, and data collection method for the Mind-Full Nepal study (Antle et al., 2015). As I will analyze the data collected from this study in this thesis, it is important to understand how the Mind-Full game works, the study design, and data collected by the Mind-Full application, even though this study is not directly a part of this thesis. This will provide background information for the data analysis methodology described in Chapter 5.

4.1. Mind-Full Application

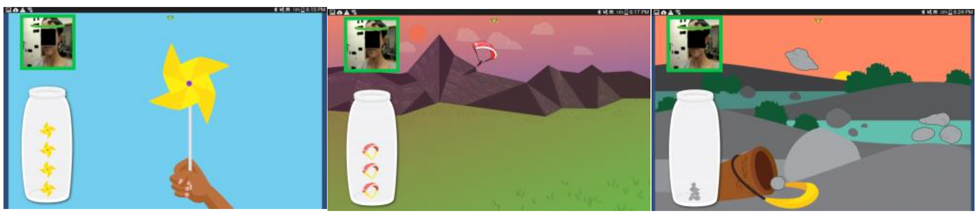


Figure 4.1. The Mind-Full games: (a) Pinwheel game; (b) Paraglider game; (c) Stones game.

Antle et al. (2015) designed an NF training tablet application called "Mind-Full" to help children who have suffered trauma to learn and practice self-regulation of anxiety (through relaxation) and attention. The Mind-Full application contains games that were designed based on familiar, everyday activities in the lives of the children in this study. There are three games: two of them provide NF training on relaxation, and the third game provides NF training on attention. They are designed to be worked through in order because a calm state may aid self-regulation of attention. One design principle for these games was to have game scenarios metaphorically depict the real-life activities of these children. This provides NF in the form of visual cues and encourages the children to perform physical actions that would shift their physiology and corresponding brainwave states to help them learn how to self-regulate around relaxation/anxiety and focus/attention (Antle et al., 2015).

The first game, called pinwheel, is a NF warm-up game for relaxation. The children get a token (reward) by rotating the pinwheel. The pinwheel rotates if the children stay relaxed for the default time (five seconds) and for the default threshold of 40 (relaxation range: 0-100). Five tokens fill up a jar. The game can be played as many times as the user likes; however, in the study, each game was often stopped after achieving a single jar of tokens due to session time constraints. This game was inspired by children blowing slowly to cause pinwheels to spin. Deep breathes, like blowing, may create a calmer brain state.

The second game, called paraglider, is an NF game for sustained relaxation. The children get a token by landing the paraglider. The paraglider lands if their relaxation score (0–100) stays above the default threshold of 40 for 11 seconds (default time). If their relaxation level drops below the threshold before the eighth second, the paraglider swirls back up and the children have to bring it back down by increasing their relaxation score above the threshold value for the remaining three seconds to get a token. Five tokens fill up a jar. This game was inspired by the children's real-life experience sitting back, relaxing, and watching paragliding in the valley that they live in.

The third game, called stones, is a NF game for sustained attention. The children get a token by piling up five stones from a basket. For collecting each stone, they have to watch the stone cross the screen, and in doing so, keep their attention value (0–100) above 40 (default threshold) for eight continuous seconds (default time). Five tokens fill up a jar. If their attention level drops, the stones go back to the basket, and they need to repeat the process to get a stone.

Generally, NF training games have a calibration phase lasting up to three minutes where the headset calibrates the relaxation and attention scores based on the individual's data. However, it is difficult to make young children with trauma disorders sit still for three minutes. To address this, the calibration was done using a networked but separate application which could be used to change default values as the children play the game. For all three games, default time (called min_time) and threshold (called base_threshold) could be changed to calibrate for individual differences and/or to increase/decrease the difficulty level in real-time. For example, once a child could finish the pinwheel game in 20 seconds with a default time of four seconds/per token, the time could be increased to eight seconds and/or the threshold can be increased to 60 to make it harder (more relaxed state required). Similarly, both the time and threshold could be decreased to reduce the difficulty of the game if a child is consistently having difficulty completing a task within default time. This calibration is done by the therapists/teachers who would connect their tablet to the child's game tablet and change the threshold and time based on their performance in real-time.

4.2. Headset Connectivity and NMM Headset Data

The NMM headset is connected to the Android tablet of the Mind-Full application using Bluetooth. When the headset gets connected to the tablet, the Unity and Java programs of the Mind-Full game connects with the headset via Bluetooth at the application level to extract headset information such as signal quality, relaxation and attention scores, and EEG amplitudes calculated by the Neurosky proprietary algorithm. The sampling rate is 1 Hz. These data are stored in the tablet in the form of log files.

The headset provides a signal quality parameter, "Poor Signal" (PS) ranging from 0–200. When PS is 0, there is no noise in data; when *P*S > 0, there is some noise; and when PS = 200, the headset is not connected. When there is a good signal quality, the UI in the game shows a green band around the picture of the user, indicating a good quality signal. When there is noise in data (PS> 0), then the display of the Mind-Full game shows a yellow or red band around the picture of the user. When the yellow or red band persists for a longer duration, the teacher/counselor would troubleshoot the connection problem and relaunch the games.

The description of the data is provided by the Neurosky developer document (Masasomeha, 2017). The relaxation and attention score range is 0 to 100. Relaxation scores below 40 indicates a stressed/anxious state of mind. Attention scores below 40 indicates a non-attentive state of mind. The Neurosky proprietary algorithm "ThinkGear" calculates the relative amplitude of eight potential brainwaves: delta (0.5 - 2.75 Hz), theta (3.5 - 6.75 Hz), low-alpha (7.5 - 9.25Hz), high-alpha (10 - 11.75 Hz), low-beta (13 - 16.75 Hz), high-beta (18 - 29.75 Hz), low-gamma (31 - 39.75 Hz), and mid-gamma (41 - 49.75 Hz). Unlike amplitudes calculated from the standard raw signal (whose unit is mV), the EEG amplitudes calculated by the ThinkGear protocol approximately ranges from 10,000 – 6M and does not have any units. This data cannot be compared with EEG amplitudes derived from other sources. For research purpose, these EEG amplitudes can be considered when the noise parameter does not signal presence of noise (Liang, Neurosky email conversation, September 2016). The raw signal ranges from -255 to 255 and was collected at one sample per second. Eye blink strength was not collected.

4.3. Nepal Study Methodology

In this study, a deductive confirmatory experiment with a waitlist control group was conducted to see if the children could 1) learn to self-regulate anxiety and attention playing Mind-Full games (learn to self-regulate), 2) complete the Mind-Full intervention (viability) and in doing this, 3) improve their ability to calm down and focus their attention at the school (transfer self-regulation skills), and 4) maintain these skills over a two-month follow-up period (maintenance of skills). To address self-regulation learning, viability, transfer, and maintenance measures, the study follows a single-phase concurrent embedded design. The primary outcomes, reported in Antle et al. (2015), are transfer and maintenance determined through between-group analysis of subjective behavioural measures. These were supplemented with objective log data of game performance, which addressed within-group measures of self-regulation learning and viability and are the focus of the methodological investigation in this thesis. The experiment was conducted in the field environment of the Nepal House Society Kaski school, an NGO-funded school in Pokhara, Nepal. The ethics information is presented in the Appendix B.

4.3.1. Participants

The participants were 22 girl children (age 5–11 years, M =7.3, SD = 2.2). The children live in extreme poverty and “suffer trauma because of violence or substance abuse at home, neglect, and/or parental death” (Antle et al., 2015). The children attended a school in Pokhara called Nepal House Kaski, funded and administered by an NGO called Nepal House Society. The children attend their academic classroom and therapy sessions, which include play and art sessions. The teachers from this school are trained by Western therapists. The children are very naïve about technology and video games.

These children have many behavioural issues and face difficulty in remaining calm and attentive in their classroom and playground. The children were classified into two equivalent groups based on their learning challenges, behavioural issues, and age. All children at the school are girls. Group 1 (intervention) consisted of nine children (two of original 11 left the school during the intervention) and Group 2 (waitlist) consisted of 12 children (one was added to the original group of 11 before intervention) for a total of 21 participants.

4.3.2. Procedure

A pre-assessment of children's calmness and attention was conducted. This pre-assessment comprised of questionnaires to obtain both quantitative and qualitative feedback from the teacher and counselors in the school. The study protocol was to complete at least 20 sessions. Group 1 was fixed as the training group, which played all three games with each session. Group 2 remained as the waitlist control. Both training and control groups received behavioural therapies that were usually taught in the school, for seven weeks. At the end of seven weeks, both groups were assessed in a post-test with the same questionnaire. For ethical reasons and to further validate the use of tablet games in improving self-regulation, Group 2 was also given training for six weeks while Group 1 did nothing different. At the end of seven weeks, a follow-up assessment was done for both groups using the same quantitative and qualitative questionnaires. It is important to understand that during the study, the children were assessed three times, (1) before the start of the training, (2) when Group 1 completed the Mind-Full training, and (3) when Group 2 completed the Mind-Full training. Log data was collected during each group's intervention.

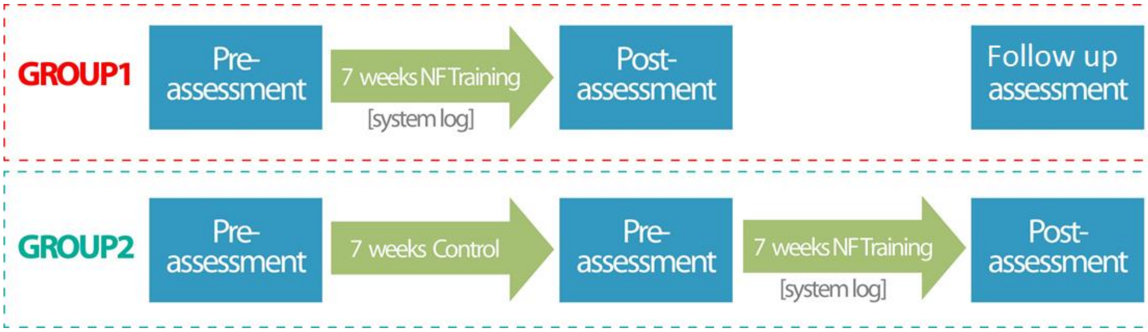


Figure 4.2. The Mind-Full study methodology.

4.3.3. Data Collection

The mixed-method approach was followed at data collection level**,** where the quantitative data was collected in the form of system log and close-ended questions (rating 0–4), and the qualitative data was collected as open-ended survey and interview questions.

The system log comprises of two types of data files. The first data file (with the filename format - "ChildtID_SessionID_Date.csv") consists of game information. For each game log file, another log file is created that consists of data from the EEG headset (with the format "ChildID_SessionID_Date_headset.csv").

The game events are created when a token is generated, when a game starts or when a game ends. Two types of events are recorded in the first file:

(1) Timestamp, game type, game started (or) game ended

(2) Timestamp, game type, Base_Threshold, Min_Time, token_number.

Here, the parameters base_threshold and min_time can be dynamically changed in the calibration tablet to vary the difficulty level of the Mind-Full games during the game-play. The ParticipantID_SessionID_Date_headset.csv records data is given by the EEG headset at the rate of 1 Hz. The file consists of data {such as Timestamp, Poor_signal_Value, relaxation score, attention score, EEG relative amplitudes, Raw signal}.

4.3.4. What needs to be done?

The between-group analysis of questionnaire data provided a subjective evidence that the treatment group had significant improvement in behaviours related to their ability to self-regulate anxiety (remaining calm) and attention during various activities in their classroom and playground at the post-test point after they had completed the intervention (Antle et al.; 2015). The log data collected from this study had further information regarding children's relaxation, attention, game performance, and EEG information. Analyzing these data might add an objective evidence to understand if the children improved their ability to relax and attend over the course of their intervention (the Mind-Full training) or if a different pattern of learning occurred. In the next chapter, I will discuss in detail the different DVs measured through the log data that correspond to relaxation, attention, and game performance and the analysis procedure that I used.

Chapter 5. Methodology: Log Data Analysis

The overall goal of this thesis is to develop and apply data analysis techniques to investigate if there is an improvement of relaxation and attention measures across sessions as recorded in the log data collected from the Nepal study's Mind-Full training. This will help us to understand if/how the Mind-Full app has helped children to learn to self-regulate calmness and attention during Mind-Full gameplay sessions. In this chapter, I describe each research question, discuss data retrieval and consolidation procedures, devise a mechanism to handle data constraints (such as missing sessions, noise, outliers), discuss the analysis I adopted, and provide the rationale for each analysis technique used. The scope of this data analysis procedure is to understand if the children improved their calmness and attention by playing the game across the sessions in their intervention.

(RQ1) What are the appropriate variables and statistical methods used to analyze the log data from the Mind-Full Nepal study in order to understand the patterns and trends in the data over several sessions? [Methodological question].

To answer this overarching RQ, I break it down into the following RQs:

(RQ 1.1) What are the dependent variables that we can calculate from our log data to answer our RQ 2?

(RQ 1.2) How to handle missing data and outliers??

(RQ 1.3) How many sessions to include in the analysis?

(RQ 1.4) What is an appropriate statistical procedure used in existing NF studies that can be applied for across-sessions analysis?

After discussing these sub-research questions, I consider the different measures that can be used from the log data. I derive a sub-research question for each dependent measure from the log data. Identifying the pattern of each dependent variable, pertaining to relaxation and attention, will help me in answering my second main research question:

(RQ2) Did the children's ability to self-regulate their relaxation and attention change (improve) across their Mind-Full training sessions? [Implication question]

To choose the measures for relaxation and attention from the log data collected from the Mind-Full study, the overarching (RQ2) is further divided into simple RQs (RQ 2.1 to 2.5), one for each of the dependent variables of noise, relaxation and attention indices, EEG amplitudes, and game performance that were present in the log data. First, I removed the influence of noise from the log data. After removing the noise, I calculated each of the dependent variables from the log data. After calculating the dependent variables, I substituted missing data and outliers for each of these dependent variables with appropriate mean values. After this, I sampled those sessions that I decided to consider for analysis. Then, I analyzed the data using statistical methods to investigate changes in the DVs across the sessions. The analysis was done separately for pinwheel, paraglider and stones game, as the goal and the nature of these games were different. These steps are explained in detail in the forthcoming section.

In this chapter, I discuss the above-mentioned methods in the following order: I first present my overall rationale for handling noise, missing data, and outliers that are common to all RQs. Then, from section 5.2 to 5.5, I present each of the dependent variables of noise, relaxation and attention indices, EEG amplitudes and game performance that will answers our main question (RQ 2). Under these sections, I present the rationale of choice for each of these DVs, how I substituted the missing data for that dependent measures, and the analysis procedure. Finally, I explain the rationale behind the statistical procedure used to answer the RQs.

5.1. Data Pre-processing

(RQ 1.1) What are the dependent variables that we can calculate from our log data to answer our RQ 2?

The raw data had two csv files per session per participant that stored game data and headset data separately. However, in the headset data log file, there were no attributes to associate the data with the game. Therefore, to identify the data corresponding to each game, I consolidated the two csv files for each session per participant using the script consolidate.py. After this, I consolidated the combined files of all the sessions to get a consolidated participant record. In this file, I labeled each data record with session number and game type for filtering data. In this way, I was able to consolidate the files for all participants where each row was represented as:

(1) {Timestamp, Session, ParticipantID, Poor_signal, relaxation score, attention score, EEG relative amplitudes, Raw signal, GameType}

If a token was collected, then the token information was appended to the format (1) as follows:

(2) {Timestamp, Session, ParticipantID, Poor_signal_Value, relaxation score, attention score, EEG amplitudes calculated by the NMM, Raw signal, GameType, TokenCount, Base_Threshold, Base_Time}.

From this file, after removing the noise, I extracted different dependent measures such as Good Quality Signal Percentage (GQSP), mean relaxation index per session, mean attention per session, mean amplitudes of brainwaves per session, and Time Taken to collect five tokens (TT) in a session. As discussed in section 2.5, the rationale for choosing these dependent variables were grounded based on the dependent variables used in the previous NF studies. The procedure for calculating each of these dependent variables are discussed in the sections 5.2 to 5.5.

5.1.1. Handling Noise, Missing Data and Outliers.

(RQ 1.2) How to handle noise, missing data, and outliers?

Noise

According to Neurosky, when the Poor_Signal = 0, the signal quality is good. When the Poor_Signal > 0, the signal quality is bad. In the log data, it was observed that if the signal quality is bad, the relaxation and attention values are either duplicated (of the previous value when Poor_Signal = 0) or set to zero. Therefore, it is unwise to include these relaxation and attention values in the data analysis and calculate DVs. I inject a data only when the Poor_Signal = 0 for mean DV calculations. The amount of noise excluded from the data is handled by RQ 2.1, which is discussed later in this section.

Missing Data

The Mind-Full training was given to two groups of children - group G1 (*N* = 9) and group G2 (*N* = 12) The study had 21 participants altogether. Two participants, P1_4 (from G1) and P2_7 (from G2) were excluded from the data analysis as they had less than 15 sessions. Therefore, *N* = 19. In this way, I accommodated all the participants who had at least 20 sessions. This method was also followed by Lim et al. (2012), Dempster et al. (2009) and Takahashi et al., (2014) where they excluded participants who quit the study. Considering the small sample size, I included all the participants for analysis who had above 20 sessions, unlike deBeus et al. (2011) and Lubar et al. (1995), who just considered good performers for analysis. Even though it is possible that there could be non-responders in the sample, I decided to include all the children who completed at least 20 sessions instead of only those children who performed well. This will be my first step of analysis for the Mind-Full study. This inclusion criteria will help me to understand if there is a gradual improvement for most of my sample size, instead of just considering only those who performed well. Since there are three different types of games with different natures and goals, the analysis was done separately for each game.

Out of 19 children, 15 children had at least 23 sessions (varying between 23 to 28 sessions). Out of the remaining four children, three had only 20 sessions, and one had 21 sessions. Even though some children had more than 23 sessions, there were around three children with only 20 Sessions and one child with only 21 sessions. Based on this summary, I conclude that all the children who did not leave the school, were able to complete the intervention (viability). I removed the first session S1 from analysis, as it was more of a training session to help the children understand the game-play. For this thesis, I will investigate the pattern for 22 sessions (session 2 to session 23). A similar method of sampling participants and sessions to include most of the data for the analysis were followed in previous research, where participants who quit the study in half were excluded from the analysis (Lee, 2009) and participants, in-spite of having few sessions missing at the end of the study were considered for the analysis (Bakhshayesh et al., 2011; Takahashi et al., 2014). The independent variable for my research question is Sessions (from sessions S2 to S23). Any missing data that is less than 10% is considered as low missing rate and tends to have very little consequence on the statistical analysis and thus does not cause a significant deviation to the answer for the research analysis (Dong &

Peng, 2013; Roth, 2002). So, for the data that I considered for analysis, I checked if the missing data was less than 10%.

Outliers

Previous works have excluded extreme outliers before performing analysis, but there were no explicit methods mentioned (Drechsler et al., 2007; Lagopoulos et al., 2009). ANOVA is sensitive to outliers and boxplot is useful in detecting outliers[27]. Osborne et al. (2004) have argued that a researcher should not just exclude outliers especially when they are not due to instrumental/measurement error but due to the legitimate occurrence, and might provide a significant contribution to the results. In these cases, they recommended to use intuition and reason out why the outlier should (or should not) be removed. For my data, I am not sure if the extreme value is due to the participant's state of mind or some other error. Since I already excluded noisy data before this step, the extreme outlier could be a legitimate reading. Therefore, I used a two-step process to exclude extreme outliers based on both session and participant. If a data point for a particular participant was identified beyond the 1.5 times the inter quartile range (IQR) in a session (Navidi, 2006), I compared this data point to other sessions to see if the participant had values in a similar range. If other sessions had data points in a similar range, then the outlying values were included for analysis; but if not, then the data point was excluded from analysis. In this way, I did not exclude the participant's effect from the analysis.

For example, Figure 5.1 (below) shows the individual participant plot with *X* axis representing every session and every participant. The *X* axis is the session, the *Y* axis is the dependent variable, and the colours represent each participant in the box plots. Box plot with 1.5 IQR was plotted for each session to identify outlying data points. Boxplot gives the outlier for each session for all participants. By the coloured trend lines, it could be understood if the was outlying data point for a session was extreme to that participant as well or if the participant had data in a similar range in other sessions. If the participant did not have a similar range of extreme values for other sessions, the outlier was eliminated. If an outlier was extreme to that session but the participant had a similar range

[27] http://www.physics.csbsju.edu/stats/box2.html

of data in other sessions, then the outlier was not eliminated. The same method was used for all research questions except for RQ 5. The RQ 5 outlier exclusion method is discussed later in section 5.6.2. The number of outliers excluded for each of the dependent variables are mentioned in chapter 6.

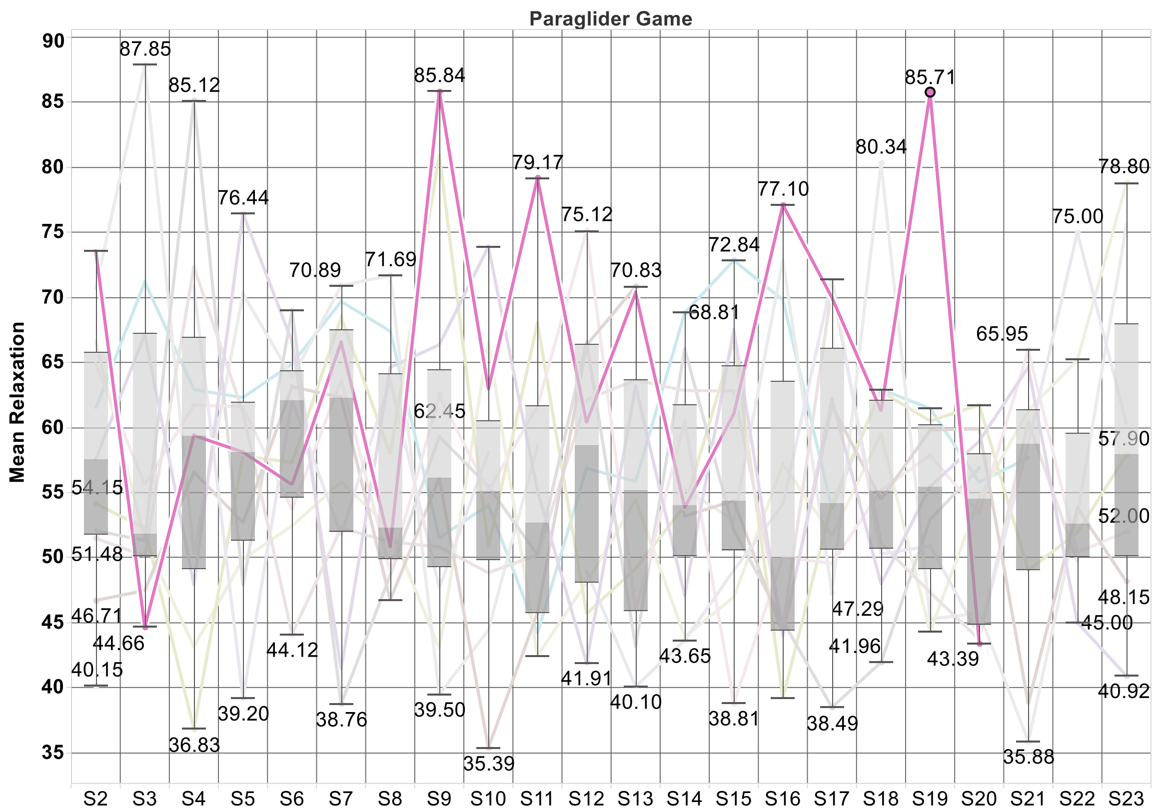


Figure 5.1. Outlier detection graph. X axis represent the session and Y axis is the dependent measure. Colour represents a participant

Missing Data Substitution

There are several methods to substitute missing data such as pair-wise or list-wise deletion, substitution with mean, median, dummy and constant values, regression method, and multiple imputation, among others. Mean substitution is the most common method used to substitute a missing value, by substituting it with the 'series mean' or 'mean of neighbouring data points'[28]. I preferred not to go with pair-wise or list-wise deletion as they

[28] https://www.ibm.com/support/knowledgecenter/SSLVMB_21.0.0/com.ibm.spss.statistics.help/replace_missing_values_estimation_methods.htm

would result in loss of other data In multiple imputation method, the missing data should be at random – however, in my case, a few participants had their last sessions missing, which is not an random occurrence (Sterne et al., 2009). I narrowed down my approach to use mean substitution as my preferred method for substituting missing data. Though previous studies have not explicitly discussed how they handled missing data, Drechsler et al. (2007) mentioned that they substituted missing data with group mean. Since I were doing repeated measurement analysis, I assumed that it would be wiser to substitute mean of neighbouring data points due to the possibility of high individual differences. This was decided as the DVs were mostly children's performance based on relaxation and attention, and I assumed that children's mindset could be relatively close to the neighboring sessions when days compared to their overall mindset during the Mind-Full training. However, for DVs such as noise occurrence, which is not based on children's mindset, I used a different type of mean substitution. More details on the method used for missing data substitution are discussed for each of the dependent variable in their individual sub-sections in Section 5.2.

5.1.2. Sampling Sessions for Analysis

(RQ 1.3) How many sessions to include in the analysis?

As discussed in Section 5.1.1, I considered 22 sessions for analysis (Session 2 to Session 23). There are several ways to sample these 22 sessions. Some of the previous studies considered several singular sessions (Gevensleben et al., 2014; Lim et al., 2012; Thomas et al., 2013) or a block of sessions (Bakhshayesh et al., 2011; Drechsler et al., 2007). Some studies have considered continuous sessions as IVs. For example, both Hillard et al. (2013) and Takahashi et al. (2014) had results that showed significant improvement in mean amplitude across sessions with 12 and 16 continuous sessions respectively. Therefore, in this analysis, I expect to see a gradual increase in children's ability to self-regulate their relaxation and attention when I include most of the continuous sessions instead of just a few blocks. Significant results in such a scenario would allow me to make stronger claims. There are certain missing sessions in-between due to connectivity issues, children being absent or only playing 2/3 games for unknown reasons, etc. in each game. These constraints are handled in a different way for each of the research questions and are discussed in their own sections.

5.2. Signal Quality across 22 sessions

(RQ 2.1) Did the percentage of Good Quality Signal Percentage (GQSP) differ across sessions for each of the Mind-Full games (pinwheel, paraglider, and stones)?

E: The noise occurrence is expected to be random for all participants and across all sessions.

In Chapter 3, I observed that the signal quality was good most of the time (95%) for the data collected under laboratory conditions; however, the participants were adults, who were instructed to stay seated until the end of the experiment. In the case of the Nepal Mind-Full study, the participants were children below the age of 12, and they showed strong symptoms of anxiety and attention issues from their behavioural pre-assessment (Antle et al., 2015). The study was conducted in a field environment. Even though the counsellors worked with them to stay still, it is possible that there could have been head movements by the participants, which can deteriorate the signal quality of the headset. Additionally, there may be interference from other signals such as Bluetooth signals, nearby wireless signals, and so on. These factors could have introduced noise into the data. For RQ2.1, I hypothesized that the noise occurrence would be random and not significant. This is because the noise in data occurs due to poor headset connectivity and/or rapid head movement of the children, which are random acts. For each game, the analysis was done separately as the nature of the games were different. Assumptions were validated before running inferential statistics. For each of the three games, the inferential statistics were run separately.

5.2.1. Dependent Variable Calculation

In the log data, the headset quality is assessed by the attribute POOR_SIGNAL, which is given by the headset at the rate of 1 Hz. POOR_SIGNAL = 0 shows that the signal quality is good. POOR_SIGNAL > 0 informs us that there is noise in the data

(Masasomeha, 2017). The percentage of good quality signal was calculated for each participant per session per game as:

Good Quality Signal Percentage (GQSP) per participant per session per game = (no. of seconds when POOR_SIGNAL = 0) / total number of seconds in that session for that participant in that game.

The GQSP was calculated using the script GQSP.java. After calculating the GQSP for each session, I excluded the outliers using the method described in section 5.1.1.

5.2.2. Handling Missing Sessions

For missing sessions, the mean of the GQSP of that participant are substituted. This substitution is influenced by Drechsler et al. (2007) where the mean of the participant group was substituted for a missing session. However, I chose to substitute the mean of the respective participant's sessions' value, as the noise could be influenced by the movement of the children because some children could stay still and others would keep moving.

Table 5.1. Analysis table for RQ 2.1.

Dependent Variable	Independent Variable	Procedure	Data Analysis
GQSP	Sessions (S2 to S23)	Remove outlier. Substitute missing sessions with mean of all sessions of a participant. Check for assumptions.	Friedman test based on assumptions violation.

Note. Analysis done separately for each game.

5.3. Change in Mean Relaxation and Attention Index Across Sessions

(RQ 2.2.R) Do both groups (G1 and G2) of children's mean relaxation scores (from NMM) improve over the course of MF intervention's pinwheel and paraglider games?

(RQ 2.2.A) Do both groups (G1 and G2) of children's mean attention scores (from NMM) improve over the course of MF intervention's stones game?

Mean relaxation score per session (for pinwheel and paraglider game separately) and mean attention score per session (for stones game) are the DVs. The change in mean scores over sessions as DVs were considered in accordance with previous studies (Lee, 2009; Lim et al., 2012; Stinson & Arthur, 2013; Thomas et al., 2013). Sessions (P = 22, from S2 to S23), and groups (G1 and G2) are the IVs. The analysis was done separately for each game to see the improvement over sessions.

5.3.1. Dependent Variable Calculation

The mean relaxation score per session was calculated for the pinwheel and paraglider games (separately) as the average of individual relaxation scores collected in that session when the POOR_SIGNAL = 0. Similarly, the mean attention score per session was calculated for the stones game as the average of the individual attention score when the POOR_SIGNAL = 0. The mean relaxation and the mean attention for all the sessions for each participant were calculated using the script AverageRA.java.

5.3.2. Handling Missing Sessions

The missing data and extreme outliers were substituted with the two neighbouring sessions' (before, after) data if their mean relaxation/attention difference was less than 20. For example, if session 4 was missing for the paraglider game and the mean amplitude relaxation of session 3 was 46 and session 5 was 56, their mean was substituted for session 4. If the difference of neighbouring session was greater than 20, or if two consecutive sessions were missing, then the mean amplitude of two nearest neighbours were considered. For example, if session 19 and 20 are missing in the stones game, the mean attention score of session 17, 18, 21, and 22 were substituted for sessions 19 and 20.

Table 5.2. Analysis table for RQ 2.2.

Dependent Variable	Independent Variable (for 1 &2)	Procedure (for 1 & 2)	Data Analysis (for 1 & 2)
1. Mean relaxation per session (for pinwheel game, for paraglider game) 2.Mean attention score per session (for stones game)	Sessions (22), Groups (G1 and G2)	Remove outlier. Substitute missing sessions. Check for assumptions. Run analysis.	Mixed ANOVA

Note. Analysis done separately for each game separately.

5.4. Neurosky Proprietary Algorithms' EEG Amplitude Change

(RQ 2.3.R.1) Do both groups (G1 and G2) of children's mean high alpha (from NMM) improve over the 22 sessions of playing the Pinwheel and Paraglider games?

(RQ 2.3.R.2) Do both groups (G1 and G2) of children's mean high alpha (from NMM) improve over the 22 sessions of playing the Pinwheel and Paraglider games?

(RQ 2.3.A.1) Do both groups' (G1 and G2) of children's mean low beta amplitude (from NMM) improve over the 22 sessions of playing the Stones game?

(RQ 2.3.A.2) Do both groups (G1 and G2) of children's mean theta amplitude (from NMM) decrease over the 22 sessions of playing the Stones game?

From Section 2.2.3, it is evident that the increase in alpha amplitude (8 to 13 Hz) improves relaxation even for novice meditators (Dempster & Vernon, 2009; Legewie et al., 1969; Mikicin & Kowalczyk, 2015; Putman, 2000; Stinson & Arthur, 2013). Hence, we chose the amplitudes of low alpha (7.5 to 9.25 Hz) and high alpha (10 to 11.75 Hz) as the measures of relaxation. Similarly, the amplitude of low beta (13 to 16.75 Hz) and theta (3 to 6.75 Hz) were chosen as the measures of attention as per Section 2.2.3. Even though previous studies considered the amplitude of Low beta or SMR waves (12–15 Hz) (Kropotov et al., 2005; J. F. Lubar et al., 1995; J. O. Lubar & Lubar, 1984; Russell-Chapin et al., 2013), some studies have considered mid-beta waves (15–18 Hz) (Bakhshayesh et al., 2011; Ogrim et al., 2012) as the measure of attention. Similarly, decrease in theta amplitude was interpreted as improvement in attention (Bakhshayesh et al., 2011 ; Leins et al., 2007 ; J. F. Lubar et al., 1995 ; Ogrim et al., 2012). Groups (G1 and G2), Sessions (2 to 22) were considered as the IVs. The analysis was done separately for each game. Missing sessions and outliers were substituted using the method explained for RQ 2.2.

5.4.1. Dependent Variable Calculation

We know that the amplitudes of each of the brainwave frequencies were given at the rate of 1 Hz by the NMM Proprietary algorithm (Masasomeha, 2017). The mean amplitudes for theta, low beta, low alpha, and high alpha were calculated for each session. The amplitudes were calculated as the average of individual amplitude per session when POOR_SIGNAL = 0. The mean EEG amplitudes were calculated for all the sessions for each participant using the script AverageEEG.java.

Table 5.3. Analysis table for RQ 2.3.

Dependent Variable	Independent Variable (for 1 & 2)	Procedure (for 1 & 2)	Data Analysis (for 1 & 2)
1.Mean low alpha amplitude and high alpha amplitude per session (for pinwheel game, for paraglider game) 2.Mean low beta amplitude and theta amplitude score per session (for stones game)	Sessions (22), Groups (G1 and G2)	Remove outlier. Substitute missing sessions. Check for assumptions.	Mixed ANOVA with sessions (S2 to S23) and groups (G1 and G2)

Note. Analysis done separately for each game.

5.5. Correlation between Neurosky Calculated Relaxation/Attention Scores and Neurosky calculated EEG Amplitudes.

(RQ 2.4.R.1) Is there a significant positive correlation between the mean relaxation score with the mean low alpha amplitude for each session in the pinwheel and paraglider games for all the participants?

(RQ 2.4.R.2) Is there a significant positive correlation between the mean relaxation score with the mean high alpha amplitude for each session in the pinwheel and paraglider games for all the participants?

(RQ 2.4.A.1) Is there a significant positive correlation between the mean relaxation score with the mean low beta amplitude for each session in the stones game for all the participants?

(RQ 2.4.A.2) Is there a significant negative correlation between the mean relaxation score with the mean theta amplitude for each session in the stones game for all the participants?

The aim of this research question is to understand if two relaxation measures, the mean relaxation score (RQ 2.2.R) and alpha amplitudes (RQ 2.3.R.1 and RQ 2.3.R.2), correlate. Similarly, to understand if two attention measures – attention score (RQ 2.2.A) and theta, and low beta amplitudes (RQ 2.3.A.1 and RQ 2.3.A.2) – correlate. This research question would help me to understand if my dependent measures of relaxation and attention from the first two research questions complement each other. It is also important to note that in an ideal situation, this research question can be answered accurately by correlating relaxation/attention scores with the amplitude of brainwave frequencies derived from raw data (collected at 512 Hz). However, in this case, the raw data was not collected. Hence, I am correlating the Neurosky calculated amplitude with the mental scores.

5.5.1. Dependent Variables

The analysis was done separately for each game. Table 5.4 (below) depicts the variables correlated for each game.

Table 5.4. Analysis table for RQ 2.4.

Game Type	Variable1	Variable2	Procedure (for 1,2,3,4)	Data Analysis (for 1,2,3,4)
1.For pinwheel and paraglider games,	Mean relaxation score	Mean amplitude of low alpha	Consider dataset that was used for data analysis of RQ 2.2 and 2.3.	If linear, Pearson If monotonic, Spearman
2. For pinwheel and paraglider game,	Mean relaxation score	Mean amplitude of high alpha	Exclude instances with missing data. Remove outliers.	If both were not present -
3.Stones game	Mean attention score	Mean amplitude of low beta	Plot scatterplot. Look for linear or monotonic relationship.	Kendal Tau b
4.Stones game	Mean attention score	Mean amplitude of theta		

Note. Analysis done separately for each game.

5.5.2. Dataset Pre-processing for Correlation

Unlike other RQs, here, the missing data was excluded from analysis. I did this because I had a sufficient set of data points to identify correlation patterns and the missing data was insignificant compared to the number of points in the data set (less than 10%). I need to detect outliers in the scatterplot relationship before correlating the dependent measures. Before correlating the dependent measures of RQ 2.2 and 2.3, extreme outliers were removed. This technique to handle outliers is different from those used in other RQs. Figure 5.2 (below) depicts how the outliers were detected before correlating mean relaxation score with the mean low alpha score for the pinwheel game. Each dependent measure (for example, mean attention and mean low beta) was plotted against each other in a scatterplot. Box plot (with 1.5 times the IQR) merged with the *X* and *Y* axis of the scatterplot to detect the outliers for each variable. If a data point fell beyond the box plot, it was considered an outlier and removed from the dataset.

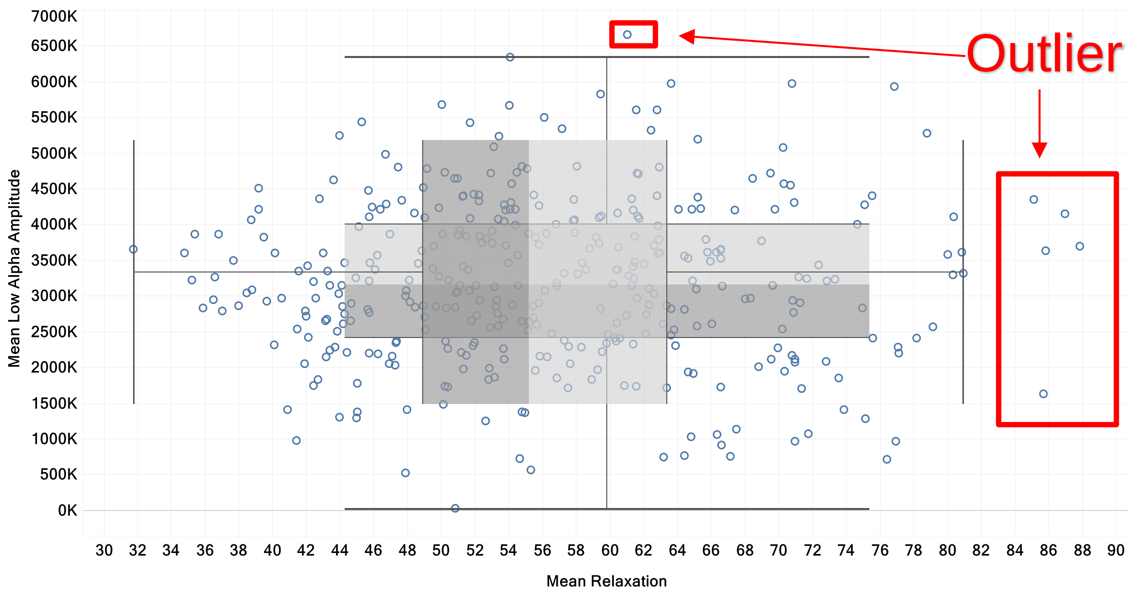


Figure 5.2. Scatter plot for mean low alpha amplitude vs mean relaxation score. Box plot along X and Y axis to detect outliers

5.6. Change in Time Taken to collect Five Tokens

(RQ 2.5) Did the time taken (in seconds) by the children to get five tokens in each session significantly reduce over 22 sessions (from session S2 to S23) for both groups of children (G1 and G2) for all three games (pinwheel, paraglider, stones) of the Mind-Full intervention?

In each session of the Nepal study, the children played the game to fill the token jar, which can hold up to five tokens. In some sessions, children collected more than five tokens. However, this was rarely the case. The protocol was for children to collect five tokens in each game. For more than 95% of the sessions, the counsellors stopped the game after collecting one jar of five of tokens according to protocol. Based on this, I decided to consider time taken to collect five tokens as the DV. The IV is the sessions (S2 to S23), and groups (G1 and G2). The analysis was done for each game separately.

Though I have two independent measures (sessions and group), there were other parameters that influenced the dependent variables. When I revisit the token collection method in Section 4.1, I can see that the time taken to collect a token is dependent on two parameters set in the calibration application by the counsellors for each game. These parameters were min_time and base_threshold. It is also important to note that there are up to five tokens that are collected in every session of a game, and each of these tokens holds a possibility to have different min_time and base_threshold. Appendix A has the individual plots of time taken to collect each token in a session, along with the adjustable parameters set for them.

Unlike previous RQs, adding or averaging the time taken to collect tokens per session might be affected by a confounding factor of varying difficulty level. Therefore, I can't run a Mixed ANOVA directly on TT. I need to consider an analysis that can accommodate the variability due to the changing min_time, threshold, and token count within a session. Vollebregt et al. (2014) conducted a similar NF study, where their dependent measure of beta and theta amplitude varied based on individual attributes such as medication, their results on neuropsychological test and so on. In this study, these parameters were treated as covariates and repeated measures ANCOVA was conducted. However, in our case, the covariates' influence is not at the participant level, but at the session and token level (each token could have a different base_threshold and min_time). Thus, I decided to normalize TT to default min_time and base_threshold before conducting repeated measures ANOVA. In this way, I can understand if the adaptability of the children to get a token improved over time and eliminate most of the variability due to min_time and base_threshold. Therefore, repeated measures ANOVA (after normalizing DV) was preferred over the repeated measures ANCOVA.

5.6.1. Normalizing Game Parameters

As mentioned in Section 4.1, the default threshold for all three games was 40. The default times were 5, 11, and 8 seconds for pinwheel, paraglider, and stones games respectively. Almost 99% of the data points had 40 as their default threshold. However, the min_time was different. Therefore, I tried to understand the relationship between min_time and time taken to collect a token. For all three games, there was a significant positive weak to moderate linear relationship between the minimum time and time taken to collect a token ($R = 0.15$ *to* 0.4, $p < 0.05$). Therefore, I considered a linear transformation of the time taken to collect a token to their respective default time and threshold. For this research question, I assume that there is a linear relationship between the time taken to collect each token and their corresponding min_time. Therefore, I normalized the time taken to collect each token by linearly transforming them based on the base_threshold and min_time set for that token (calculated using MS Excel). Normalized time taken to collect each token is calculated as: normalized time taken to collect token A = (actual time taken to collect a token A * default base_threshold * default min_time) / (calibrated base_threshold set for token A * calibrated min_time set for token A).

For example, if the actual time taken to collect a token was 13 seconds. The parameters set by the counsellors were calibrated base_threshold and calibrated min_time. Say, the calibrated base_threshold was 30 and calibrated min_time (in seconds) for which they should have a relaxation score above 30 was 11 seconds. If the default base_threshold was 40 and default min_time was five seconds, then, the normalized time taken to collect the token = (13*5*40)/ (30*11) =7.9 seconds. Once I had the normalized time taken for each token of a game, I added all the normalized time taken to collect five tokens. This would enable me to calculate the DV 'normalized time taken to collect five tokens in a session'. I call this normalized TT. However, there are certain special cases such as no tokens collected in a session or <5 tokens collected in a session. This is discussed in Section 5.6.2.

5.6.2. Intrusion of Poor Signal in the Dataset

The Mind-Full app used in the Nepal study did not exclude poor signal data from the relaxation/attention score calculation. From the log data, it was observed that when the noisy signal (Poor_Signal > 0) lasted for a shorter time duration (less than or equal to six seconds), then the headset substituted the value from the previous instance when the Poor_Signal was 0. However, this duplication of values was sometimes observed even if noisy signals lasted more than six seconds. We manually excluded such duplicate values which lasted more than six seconds.

5.6.3. Handling Missing Data

There are a few cases where less than five tokens were collected for a game in a session. After normalizing the time taken to collect a token, these special cases were handled as follows:

(1) For a game, if the session had only three or four tokens, I took the average of normalized time taken to collect first three tokens and substituted them for the fourth and fifth tokens.

(2) For a game, if the session had only one or two tokens, I considered the session to be a missing session and excluded it from the analysis.

After this, extreme outliers were removed and considered missing data. The average of normalized TT of the nearest neighbours were substituted to the missing session. If the difference between the neighbouring values were more than 5 * default_time, then the four nearest neighbours were considered for substitution. If two consecutive sessions were missing, the mean values of the four nearest neighbours were substituted. In this way, I could get a representation of the performance from the nearby sessions.

Table 5.5. Analysis table for RQ 2.5.

Independent Variables	Dependent Variable	Covariates	Procedure	Data Analysis
Sessions (S2 to S230 Groups (G1 and G2)	Time taken to collect 5 tokens TT. (normalized)	Min_time, base_threshold,	Normalize time taken to collect each token based on default value of the covariates. Remove outliers and substitute missing sessions. Check for assumptions.	Mixed ANOVA with groups (G1, G2) and session (S2 to S23)

Note: Analysis done separately for each game.

5.7. Rationale for the Statistical Analysis

(RQ 1.4) What is an appropriate statistical procedure used in the existing NF studies to observe the change in the relaxation and attention across the session?

As discussed in Chapter 4, I note that the study was designed to give Mind-Full training to the participant groups (G1 and G2) over about 20-27 sessions, depending how many they could fit into seven weeks. In Section 5.1.1, I decided to consider 22 sessions (S2 to S23) for analysis. I have two groups in the Mind-Full study design. Based on the rationale I discussed in section 2.4.5, I chose to perform 2×22 mixed ANOVA to identify the significant pattern of the dependent measures across the training period with groups (G1 and G2), and sessions (S2 to S23) as independent measures. This method was also used in NF studies with similar study designs. (Bakhshayesh et al., 2011 ; Egner et al., 2002 ; Gevensleben et al., 2014 ; Kirenskaya et al., 2011 ; Lagopoulos et al., 2009 ; McDonald, 1974)

To perform a Mixed ANOVA, I checked the assumptions such as normality, sphericity (for repeated measurements), and homogeneity (for groups). I chose the Shapiro Wilk test for testing normality, Mauchly's test for testing sphericity of variance, and Levene's test for homogeneity of variance. Since the repeated measurement (22 sessions) was greater than the sample size (19 participants), the sphericity was undefined. To address this, the results were reported with Greenhouse-Geisser correction. ANOVA is robust to handle violations for slight deviations from normality, for approximately equal sample sizes (Judd, McClelland, & Culhane, 1995). There were some deviations from the assumptions, however, I still ran Mixed ANOVA to consider both main and interaction effects. Doing this might increase the chances of type-1 error (false positive). Oberfeld (2013) suggested that the type-1 error could be controlled by applying an appropriate correction. In our case, I applied the Greenhouse-Geisser correction to address the issue with sphericity. Along with that, to make sure that the significance should not be overstated, I decided to report the main effects from the non-parametric statistics if any significant results were observed from the Mixed ANOVA results. The effect sizes were reported with partial-eta squared.

Chapter 6. Results

In this chapter, I present the results for RQ 2.1 to 2.5 in the following order: missing data and outliers, descriptive statistics, and inferential statistics for each game.

6.1. Data Pre-processing

Participants P1_4 and P2_7 were excluded as they had less than 15 sessions. Therefore, *N* = 19. All 19 children played from sessions S1 to S20 and 15 of them children had played from sessions S1 to S23. Out of these, five children had up to 27 sessions. However, I considered the sessions for analysis as S2 to S23 as most participants (15 out of 19) played these sessions. Session 1 was eliminated from the dataset, as the children were taught to use the tablet, and the counselors were consistently monitoring and changing their details.

6.2. Signal Quality Across 22 Sessions

(RQ 2.1) Did the percentage of Good Quality Signal Percentage (GQSP) differ across sessions for each of the Mind-Full games (pinwheel, paraglider, and stones)?

E: The GQSP should be a random occurrence, and the sessions should not have influence over the GQSP.

6.2.1. Pinwheel Game: GQSP

Here, the DV is GQSP. The IV is session (S2 to S23).

Missing Data and Extreme Outliers

For the pinwheel game, 15 data points were missing (including extreme outliers) out of 418 data points, thus contributing to 4% of total missing data.

Descriptive Statistics

In this section, I present the descriptive statistics of the GQSP for the pinwheel game. Tables 6.1 (below) provides the mean and standard deviation of the GQSP for 19 participants from sessions S2 to S23. Figure 6.1 shows the trend of the data for each participant.

Table 6.1. Descriptive statistics of GQSP per session for pinwheel game (*N* = 19).

Sessions	Mean	Std. Deviation
S2	71.79	28.94
S3	71.06	27.43
S4	80.13	24.61
S5	67.59	26.97
S6	77.81	26.28
S7	74.53	24.93
S8	67.18	28.88
S9	68.88	27.72
S10	74.27	23.89
S11	69.47	28.20
S12	78.63	25.15
S13	57.70	34.03
S14	73.51	22.49
S15	76.07	26.71
S16	72.94	25.35
S17	71.47	30.70
S18	77.88	28.99
S19	72.91	29.60
S20	74.33	27.73
S21	79.77	23.19
S22	80.11	24.70
S23	79.70	26.15

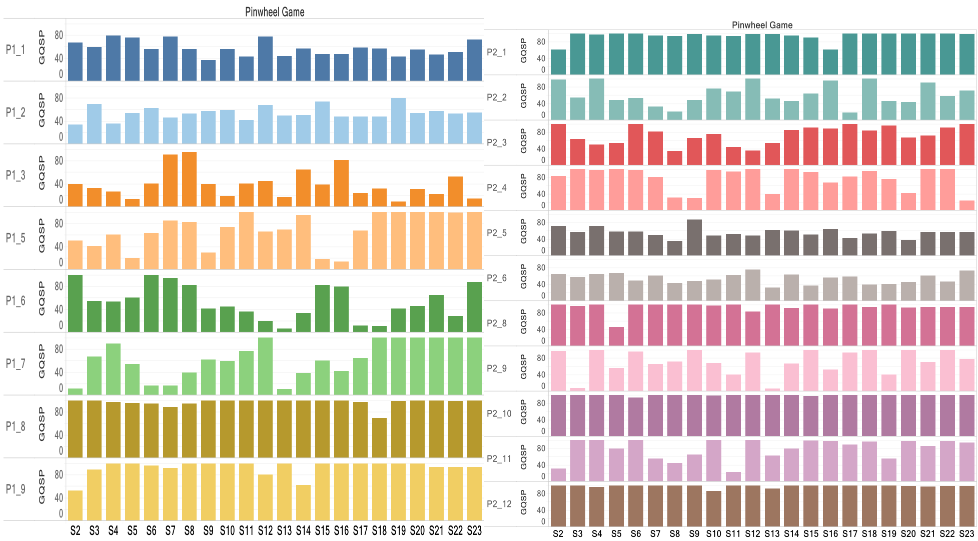


Figure 6.1. Individual plot of GQSP vs. sessions (S2 to S23) for pinwheel game (*N*=19).

Effect of Sessions on GQSP

The Friedman test revealed that there were no significant differences in the GQSP across sessions (S2 to S23) for the pinwheel sessions, $N = 19$, $\chi^2(21) = 18.22$, $p = 0.63$, and the Kendall correlation was $W = 0.04$, indicating a very weak relationship between the IV and DV. Therefore, the GQSP did not significantly change across the sessions for the pinwheel game.

6.2.2. Paraglider Game: GQSP

Here, the DV is GQSP. The IV is session (S2 to S23).

Missing Data and Extreme Outliers

For the paraglider game, 37 data points were missing and had five extreme outliers, thus yielding 10% of total missing data.

Descriptive Statistics

In this section, I present the descriptive statistics of the GQSP for the paraglider game. Tables 6.2 (below) provides the mean and standard deviation of the GQSP for 19 participants from sessions S2 to S23. Figure 6.2 shows the trend of the data for each participant.

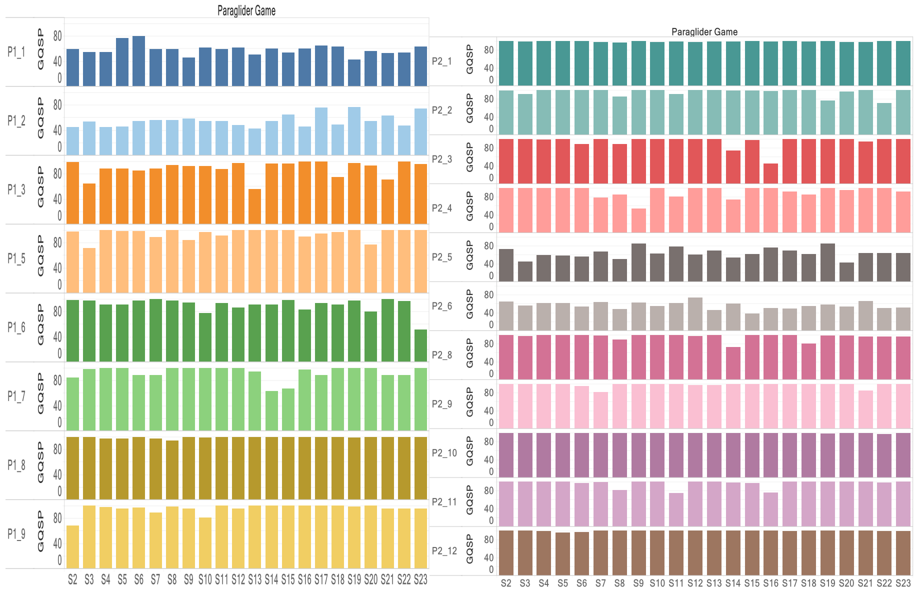


Figure 6.2 **Individual plot of GQSP vs. sessions (S2 to S23) for paraglider game (*N*=19).**

Table 6.2. **Descriptive of GQSP for paraglider game (*N* = 19).**

Sessions	Mean	Std. Deviation
S2	88.95	17.58
S3	85.70	20.47
S4	89.20	18.53
S5	90.01	16.77
S6	89.04	16.28
S7	87.12	15.30
S8	85.84	17.99
S9	88.19	18.26
S10	88.46	17.14
S11	88.15	15.34
S12	90.51	16.45
S13	86.81	21.22
S14	84.23	18.15
S15	88.12	19.81
S16	85.45	20.21
S17	90.99	14.85
S18	87.23	17.66

Sessions	Mean	Std. Deviation
S19	91.04	16.36
S20	86.69	19.65
S21	88.29	16.00
S22	87.39	19.22
S23	88.79	17.58

Effect of Sessions on GQSP

The Friedman test revealed that there were no significant changes in the GQSP from S2 to S23 for the paraglider game, $\chi^2(21) = 16.53$, $p = 0.73$, and the Kendall correlation was $W = 0.04$, indicating a very weak relationship between the IV and DV. Therefore, the GQSP did not significantly change across the sessions for the paraglider game.

6.2.3. Stones Game: GQSP

Here, the DV is GQSP. The IV is session (S2 to S23).

Missing Data and Extreme Outliers

For the stones game, 42 data points were missing (including extreme outliers). Therefore, 10% of the data were missing.

Descriptive Statistics

In this section, I present the descriptive statistics of the GQSP for the stones game. Tables 6.3 (below) provides the mean and standard deviation of the GQSP for 19 participants from sessions S2 to S23. Figure 6.3 shows the trend of the data for each participant.

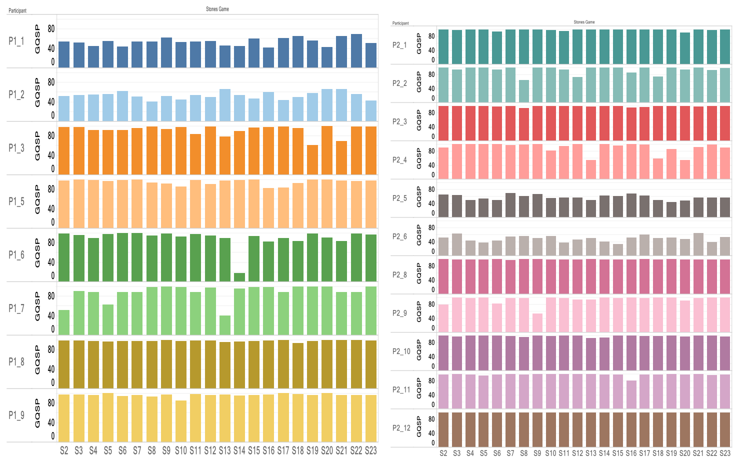


Figure 6.3. **Individual plot of GQSP vs. sessions (S2 to S23) for stones game (*N*=19).**

Table 6.3. **Descriptive statistics of GQSP for stones game (*N* = 19).**

Sessions	Mean	Std. Deviation
S2	86.04	19.87
S3	89.47	17.11
S4	87.08	21.02
S5	86.38	21.06
S6	86.35	20.27
S7	89.39	17.64
S8	86.41	19.62
S9	87.82	19.28
S10	86.59	19.19
S11	86.89	20.07
S12	86.93	19.74
S13	81.79	22.35
S14	83.60	25.76
S15	88.28	20.88
S16	86.47	18.16
S17	88.57	17.49
S18	84.66	19.64
S19	86.99	20.35
S20	85.38	21.12
S21	88.34	15.36

Sessions	Mean	Std. Deviation
S22	88.89	18.59
S23	88.45	20.08

Effects of Sessions on GQSP

The Friedman test revealed that there were no significant changes in the GQSP from S2 to S23 for the stones game, *N=19,* $\chi^2(21) = 27.85$, $p = 0.1$, and the Kendall correlation was $W = 0.07$, indicating a very weak relationship between the IV and DV. Therefore, the GQSP did not significantly change across the sessions for the stones game.

6.2.4. Summary

For all three games, the GQSP did not significantly vary across the sessions. Our expectation that the noise occurrence was random was supported.

6.3. Change in Mean Relaxation and Attention across Sessions

(RQ 2.2.R) Do both groups (G1 and G2) of children's mean relaxation scores (from NMM) improve over the course of MF intervention's pinwheel and paraglider games?

(RQ 2.2.A) Do both groups (G1 and G2) of children's mean attention scores (from NMM) improve over the course of MF intervention's stones game?

E: In the pinwheel and paraglider games, all children's mean relaxation scores will increase across the sessions. In the stones game, all children's mean attention scores will increase across the sessions. I expect to see no difference between G1 and G2.

6.3.1. Pinwheel Game: Mean Relaxation

The dependent variable for RQ 2.2.R is the mean relaxation per session. The independent variables are Sessions (S2 to S23) and Groups (G1 and G2). The analysis was done separately for the pinwheel game.

Missing Data and Extreme Outliers

For the pinwheel game, 11 data points were missing and there were three extreme outliers, thus yielding 3% missing data for the total data points collected (*N*=418)

Descriptive Statistics

In this section, I present the descriptive statistics of the mean relaxation per session for the pinwheel game. Tables 6.4 (below) provides the mean and standard deviation of the mean relaxation for 19 participants from sessions S2 to S23. Figure 6.4 shows the trend of the data for each participant.

Table 6.4. Descriptive statistics of (mean relaxation per session) for pinwheel game (*N* = 19).

Sessions	Mean	Std. Deviation
S2	54.67	14.30
S3	53.09	11.65
S4	56.82	13.61
S5	56.56	10.01
S6	56.27	10.10
S7	55.44	11.51
S8	51.73	9.48
S9	49.20	10.65
S10	51.10	11.81
S11	49.28	9.27
S12	57.78	11.74
S13	51.13	9.38
S14	54.70	9.08
S15	50.17	13.78
S16	55.98	15.38

Sessions	Mean	Std. Deviation
S17	54.42	12.01
S18	52.58	8.06
S19	54.83	11.86
S20	54.72	12.45
S21	60.08	9.64
S22	50.79	8.52
S23	57.09	12.23

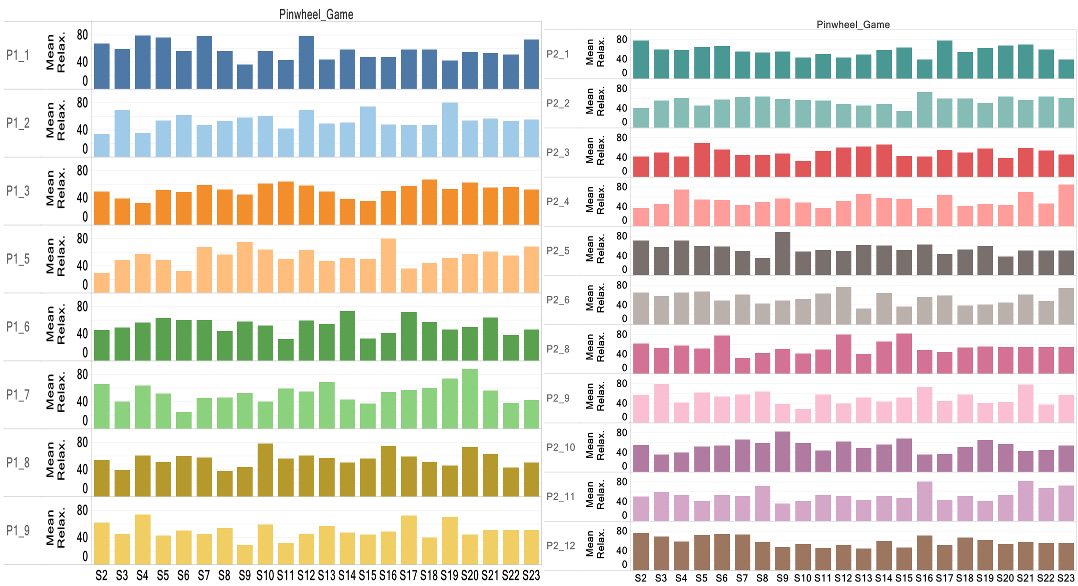


Figure 6.4. Individual plot of mean relaxation vs. sessions (S2 to S23) for pinwheel game (N = 19).

Assumptions and Analysis

The Shapiro Wilk test revealed that the mean relaxation was normally distributed from S2 to S23 (all *ps* > 0.05), except for S3 ($p = 0.001$). The Mauchly's test revealed that the sphericity of variance across the sessions (within-subject) was undefined, as sample size ($N = 19$) was less than repeated measurement counts ($N = 22$). Therefore, Greenhouse–Geisser correction was applied. The Levene's test revealed that the homogeneity of variance between groups G1 ($N = 8$) and G2 ($N = 11$) had not been violated, ($p > 0.05$). The 2 × 22 Mixed ANOVA with Greenhouse–Geisser correction was done with sessions and groups as IV and the mean relaxation per session as DV.

Effects of Sessions and Groups on Mean Relaxation

There was no significant interaction between sessions and groups over the mean relaxation score for the pinwheel game, F (9.18,156.15) = 1.2, p = 0.29, η^2 = 0.06. There was no significant change in mean relaxation across sessions for the pinwheel game, N = 19, F (9.18, 156.15) = 1.04, p = 0.4, η^2 = 0.05. There was no significance difference between G1 (N = 8) and G2 (N = 11) on the change in mean relaxation per session for the pinwheel game, F (1,17) = 0.64, p = 0.43 [p > 0.05], η^2 = 0.03

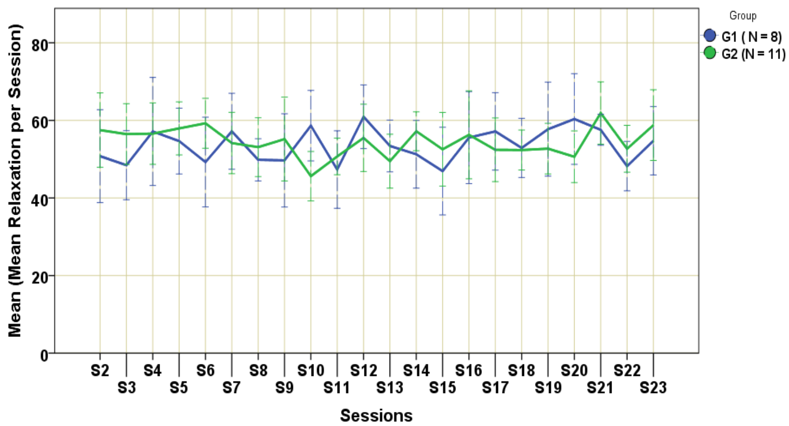


Figure 6.5. Mean (mean relaxation /session) vs. 22 sessions for G1 (N = 8) and G2 (N = 11) for pinwheel game.

6.3.2. Paraglider Game: Mean Relaxation

The dependent variable for RQ 2.2.R is the mean relaxation per session. The independent variables are Sessions (S2 to S23) and Groups (G1 and G2). This analysis of RQ 2.2.R was separately conducted for the paraglider games.

Missing Data and Extreme Outliers

For the paraglider game, 37 data points were missing and there were two extreme outliers, thus contributing to 9% of total missing data.

Descriptive Statistics

In this section, I present the descriptive statistics of the mean relaxation per session for the paraglider game. Tables 6.5 (below) provides the mean and standard deviation of the mean relaxation for 19 participants from sessions S2 to S23. Figure 6.6 shows the trend of the data for each participant.

Table 6.5. Descriptive statistics of (mean relaxation per session) for paraglider game (*N* = 19).

Sessions	Mean	Std. Deviation
S2	54.76	9.37
S3	53.43	8.20
S4	55.91	12.19
S5	56.60	10.98
S6	57.81	10.71
S7	57.97	11.19
S8	55.61	9.00
S9	59.38	13.42
S10	57.01	9.48
S11	56.16	9.02
S12	56.60	9.37
S13	54.48	9.33
S14	53.32	8.11
S15	54.67	10.38
S16	56.50	13.71
S17	56.92	12.92
S18	56.41	9.34
S19	58.47	11.50
S20	53.27	10.28
S21	56.83	8.43
S22	56.44	8.87
S23	59.40	10.58

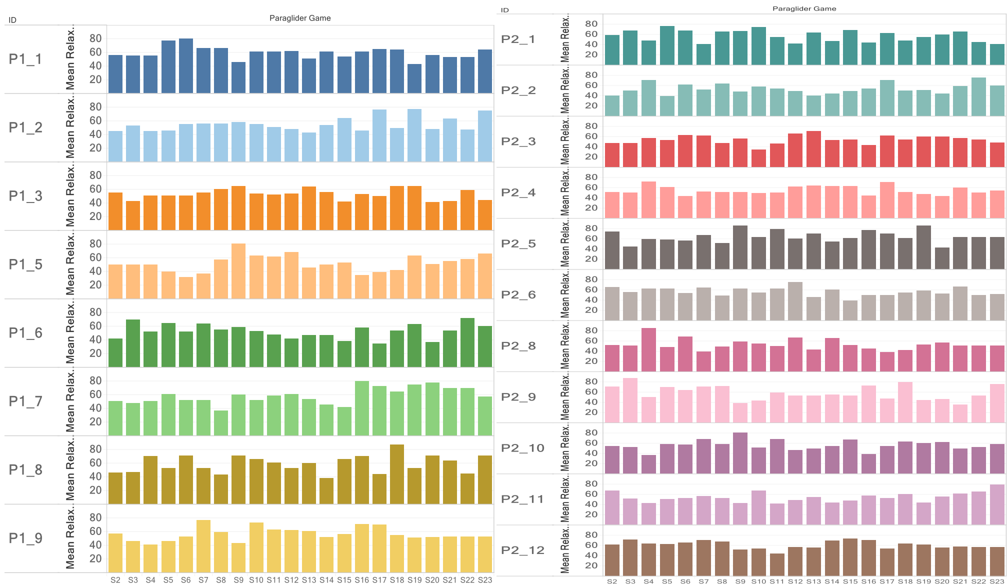


Figure 6.6. **Individual plot of mean relaxation vs. sessions (S2 to S23) for paraglider game (*N* = 19).**

Assumption and Analysis

The Shapiro Wilk test revealed that the mean relaxation was normally distributed from S2 to S23 (all *ps* > 0.05), except for S3 ($p = 0.001$). The Mauchly's test revealed that the sphericity of variance across the sessions (within-subject) was undefined, as sample size ($N = 19$) was less than repeated measurement counts ($N = 22$). Therefore, Greenhouse-Geisser correction was applied. The Levene's test revealed that the homogeneity of variance between groups G1 ($N = 8$) and G2 ($N = 11$) had not been violated, ($p > 0.05$). The 2 × 22 Mixed ANOVA with Greenhouse–Geisser correction was done with sessions and groups as IV and the mean relaxation per session as DV.

Effects of Sessions and Groups on Mean Relaxation

There was no significant interaction between groups and sessions on the mean relaxation per session for the paraglider game, $N = 19$, F (9.64, 164.03) = 0.63, $p = 0.77$, $\eta^2 = 0.03$. There was no significant change in mean relaxation across sessions for the paraglider game, $N = 19$, F (9.64, 164.03) = 0.62, $p = 0.9$, $\eta^2 = 0.03$. There was no significance difference between G1 ($N = 8$) and G2 ($N = 11$) on the change in the mean relaxation score per session for the paraglider game, F (1,17) = 0.25, $p = 0.62$, $\eta^2 = 0.01$.

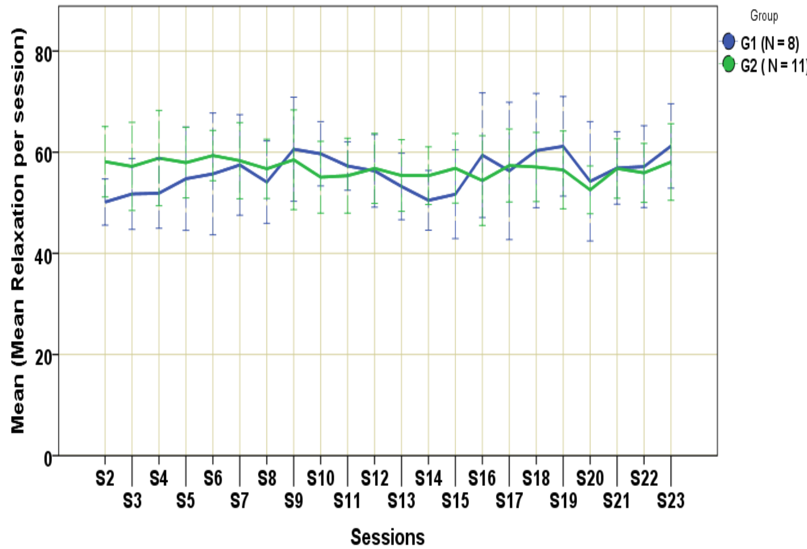


Figure 6.7. **Mean (mean relaxation /session) vs. 22 sessions for G1 (*N*= 8) and G2 (*N* = 11) for paraglider game. Error bar: 95% CI.**

6.3.3. Stones game

The dependent variable for RQ 2.2.A is the mean attention per session. The independent variables are Sessions (S2 to S23) and Groups (G1 and G2). The analysis was separately conducted for the stones game.

Missing Data and Extreme Outliers

For the stones game, 42 data points were missing out of 418 and three extreme outliers were present. Therefore, 10% of the data were missing.

Descriptive Statistics

In this section, I present the descriptive statistics of the mean attention per session for the stones game. Tables 6.6 (below) provides the mean and standard deviation of the mean attention for 19 participants from sessions S2 to S23. Figure 6.8 shows the data trend for each participant.

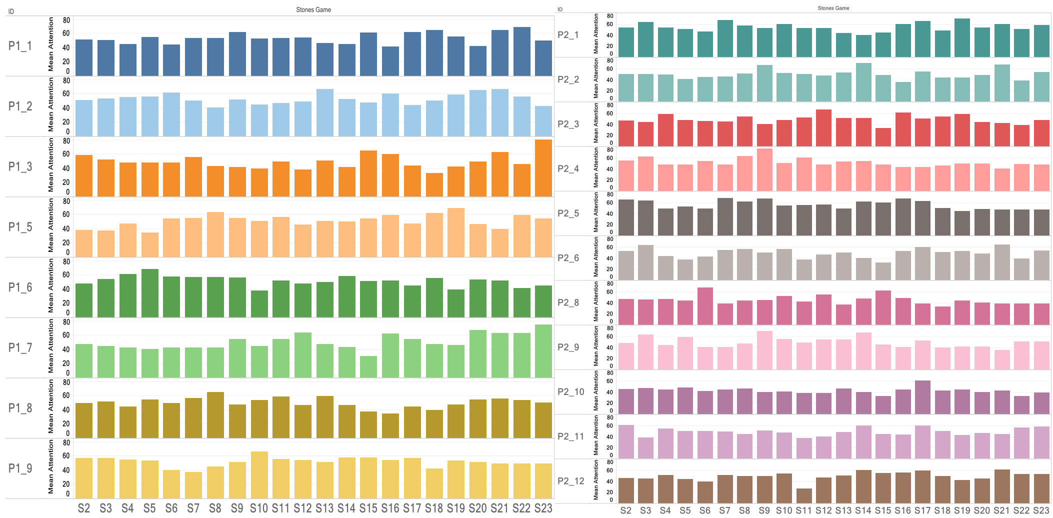


Figure 6.8. Individual plot of mean attention vs. sessions (S2 to S23) for stones game (*N* = 19).

Table 6.6. Descriptive statistics of (mean attention score per session) for stones game (*N* = 19).

Session	Mean	Std. Deviation
S2	51.33	6.34
S3	52.18	8.42
S4	49.77	5.27
S5	49.16	8.12
S6	48.60	7.65
S7	50.81	8.75
S8	51.93	7.88
S9	53.12	8.55
S10	50.74	6.91
S11	49.00	8.94
S12	50.22	7.57
S13	50.51	5.96
S14	52.11	9.40
S15	48.20	10.85
S16	51.50	9.66
S17	53.17	8.00
S18	47.47	8.34
S19	49.87	9.24

Session	Mean	Std. Deviation
S20	49.44	7.27
S21	52.65	10.90
S22	49.19	9.40
S23	51.09	7.91

Assumption and Analysis

The Shapiro Wilk test revealed that the mean attention was normally distributed from S2 to S23 (all *p*s > *0.05),* except for S19, S20, S21, and S23 (*p*s < 0.05). The Mauchly's test revealed the sphericity of variance as undefined, as sample size (*N* = 19) was less than repeated measurement counts (*N* = 22). Therefore, Greenhouse–Geisser correction was applied. The Levene's test showed that the homogeneity of variance between groups G1 (*N* = 8) and G2 (*N* = 11) was not significant, all *p* > 0.05 (except S3, S11, and S23). The 2 × 22 Mixed ANOVA with Greenhouse–Geisser correction was done with sessions and groups as IV and the mean attention score per session as DV.

Effects of Sessions and Groups on Mean Attention

There was no significant interaction between sessions and groups on the mean attention per session for the stones game, $N = 19$, F (9.81, 166.78) = 1.29, $p = 0.24$, $\eta^2 = 0.07$. There was no significant change in mean attention per session across sessions for the stones game, $N = 19$, F (9.81, 166.78) = 0.71, $p = 0.7$, $\eta^2 = 0.04$. There was no significance difference between G1 ($N = 8$) and G2 ($N = 11$) on the change in the mean attention for the stones game, F (1,17) = 1.72, $p = 0.2$, $\eta^2 = 0.09$.

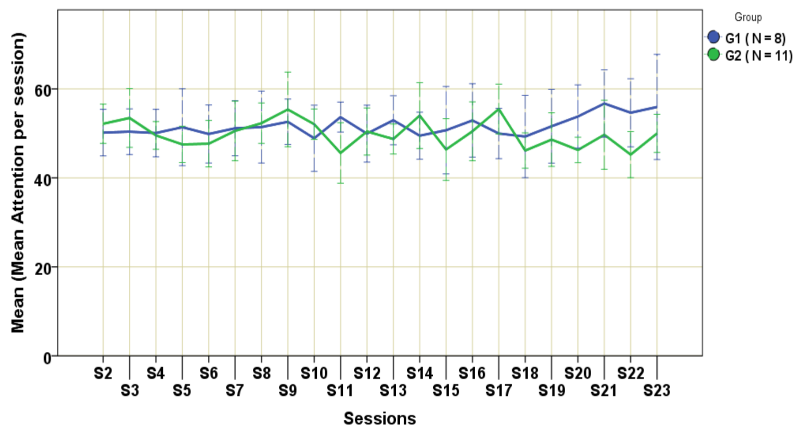


Figure 6.9. **Mean (mean attention per session) vs. 22 sessions for G1 (*N*= 8) and G2 (*N* = 11) for stones game. Error bar: 95% CI.**

6.3.4. Summary

The results did not meet my expectations with respect to change in pattern across sessions. There was no significant pattern or trend observed in the mean relaxation score during the period of Mind-Full intervention training for both the pinwheel and paraglider games. For the stones game, there was no significant trend or pattern in the mean attention score observed during the Mind-Full intervention training. As expected, the groups did not have any significant difference in the mean relaxation or mean attention across the sessions.

6.4. Change in EEG Amplitude over Sessions

(RQ 2.3.R.1) Do both groups (G1 and G2) of children's mean high alpha (from NMM) improve over the 22 sessions of playing the pinwheel and paraglider games?

(RQ 2.3.R.2) Do both groups (G1 and G2) of children's mean high alpha (from NMM) improve over the 22 sessions of playing the pinwheel and paraglider games?

(RQ 2.3.A.1) Do both groups (G1 and G2) of children's mean low beta amplitude (from NMM) improve over the 22 sessions of playing the stones game?

(RQ 2.3.A.2) Do both groups (G1 and G2) of children's mean theta amplitude (from NMM) decrease over the 22 sessions of playing the stones game?

E: In the pinwheel and paraglider game, all children's low and high alpha amplitudes will increase across the sessions. In the stones game, all children's mean low beta amplitudes will increase across the sessions and mean theta amplitude will decrease across the sessions. There will be no difference between G1 and G2.

6.4.1. Dependent and Independent Variables

For research question (RQ 2.3.R.1), the mean low alpha amplitude per session was the dependent variable. For RQ 2.3.R.2, mean high alpha amplitude per session was the dependent measure. For RQ 2.3.A.1, mean low beta amplitude was the dependent measure. For RQ 2.3.A.2, mean theta amplitude was the dependent measure. Sessions (S2 to S23) and groups (G1 and G2) were IVs. RQ 2.3.R.1 and RQ 2.3.R.2 were conducted separately for the pinwheel and paraglider games. RQ 2.3.A.1 and RQ 2.3.A.2 were conducted separately for the stones game alone.

6.4.2. Pinwheel Game: Low Alpha

For research question (RQ 2.3.R.1), the mean low alpha amplitude per session was the DV. Sessions (S2 to S23) and groups (G1 and G2) were the IVs. The result reported in this section was for the pinwheel game.

Missing Data and Extreme Outliers

For the pinwheel game, the missing data and the extreme outliers were 4% of total data (19 missing data/418 total data points).

Descriptive Statistics

In this section, I present the descriptive statistics of the mean low alpha amplitude (no units) per session for the pinwheel game. Tables 6.7 (below) provides the mean and standard deviation of the mean low alpha amplitudes for 19 participants from sessions S2 to S23. Figure 6.10 shows the data trend for each participant.

Table 6.7. Descriptive statistics of (mean low alpha amplitude per session) for pinwheel game (*N* = 19).

Sessions	Mean	Std. Deviation
S2	3254451.58	1149312.94
S3	3277479.68	1448562.10
S4	3061921.08	1227342.20
S5	2629301.82	1087822.91
S6	3120007.05	1121136.89
S7	3611190.05	1092034.44
S8	3473773.18	1186459.07
S9	3293375.89	949353.62
S10	3075791.37	1163081.45
S11	3133314.32	1079039.12
S12	3572442.84	1525774.82
S13	3504529.00	1091114.69
S14	3211127.95	1163048.85
S15	3391663.30	988978.61
S16	3396834.46	1290038.93
S17	3427488.24	1490515.17
S18	3285750.05	976811.61
S19	3338582.68	1118629.12
S20	3402947.32	947841.61
S21	3367080.06	1531383.51
S22	2987346.12	1114268.68
S23	3002380.60	1240984.64

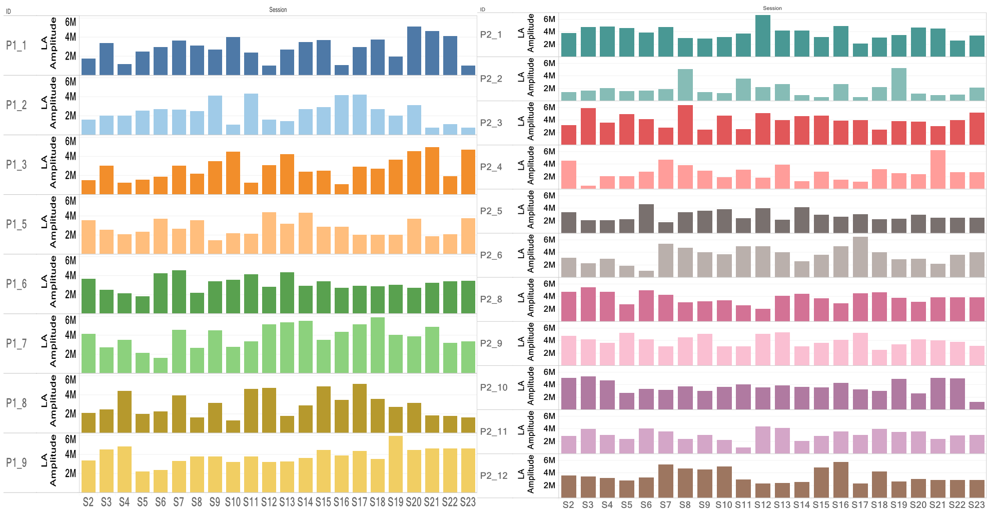


Figure 6.10. Individual plot of mean low alpha amplitude vs. sessions (S2 to S23) for pinwheel game (*N* = 19).

Assumptions and Analysis

The Shapiro Wilk test revealed that the mean low alpha amplitude was normally distributed from S2 to S23 for 19 participants, $p > 0.05$. The Mauchly's test did not give *p*-value as the sample size ($N = 19$) was less than repeated measurements (22 sessions). The Levene's test revealed that the homogeneity of variance test had not been violated for all sessions (all *p*s > 0.05), except for S2 ($p = 0.016$) and S5 ($p = 0.006$). The 2 × 22 Mixed ANOVA with Greenhouse–Geisser correction was done with groups and session as IV, and the mean low alpha amplitude per session as DV.

Effects of Sessions and Groups on Mean Low Alpha Amplitude

There was no significant interaction between sessions and groups on the mean low alpha amplitude across sessions, $N = 19$, $F\ (9.06, 154.16) = 1.17$, $p = 0.31$, $\eta^2 = 0.06$. There was no significant change in mean low alpha amplitude across sessions for the pinwheel game, $N = 19$, $F\ (9.06, 154.16) = 0.95$, $p = 0.47$, $\eta^2 = 0.05$. There was no significant difference between G1 ($N = 8$) and G2 ($N = 11$) in mean low alpha amplitude, $F\ (1,17) = 1.15$, $p = 0.29$, $\eta^2 = 0.06$.

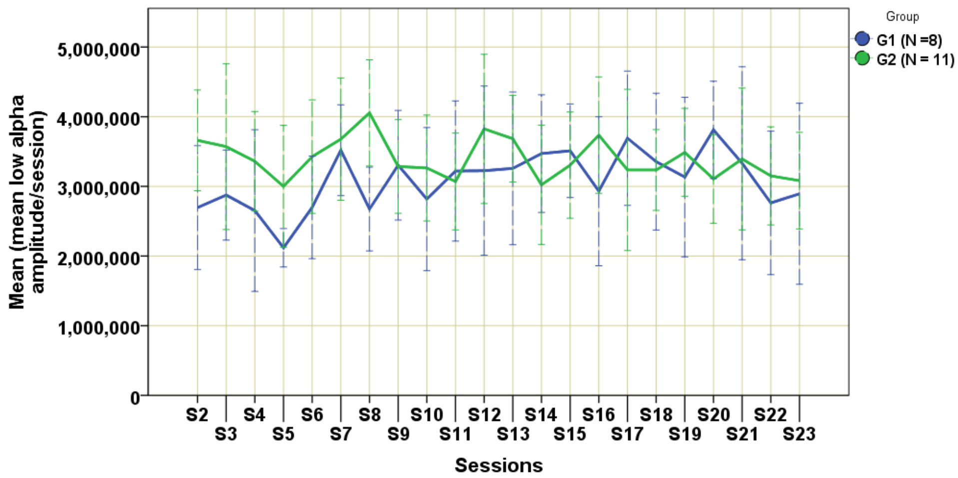


Figure 6.11. Mean (mean low alpha amplitude per session) vs. 22 sessions for G1 (*N*= 8) and G2 (*N* = 11) for pinwheel game.

6.4.3. Pinwheel Game: High Alpha

For (RQ 2.3.R.2), mean high alpha amplitude per session was the DV. Sessions (S2 to S23) and groups (G1 and G2) were the IVs. The result reported in this section was for the pinwheel game

Missing Data and Extreme Outliers

For the pinwheel game, the missing data and extreme outliers were 4% of total data (20 missing data points/418 total data points).

Descriptive Statistics

In this section, I present the descriptive statistics of the mean high alpha amplitudes (no units) per session for the pinwheel game. Tables 6.8 (below) provides the mean and standard deviation of the mean high alpha amplitudes for 19 participants from sessions S2 to S23. Figure 6.12 shows the data trend for each participant.

Table 6.8. Descriptive statistics of (mean high alpha amplitude per session) for pinwheel game (N = 19).

Sessions	Mean	Std. Deviation
S2	3092389.74	1240815.06
S3	2819382.05	1171122.99
S4	2463725.21	1561321.13
S5	2206897.34	1527652.77
S6	2914240.74	1703026.70
S7	2977940.11	1368496.48
S8	2896279.00	1386383.25
S9	2590775.00	1192356.44
S10	2390437.26	887449.68
S11	2312234.42	1070823.11
S12	2584654.84	1509605.61
S13	2183688.61	1377838.17
S14	2562204.16	1146389.26
S15	2401811.05	1138517.20
S16	2431855.32	1486246.72
S17	2267275.11	1255958.58
S18	2715894.00	1396373.71
S19	2269624.95	1087919.65
S20	2487875.78	1160546.55
S21	2046079.67	1307129.63
S22	1849006.74	1187387.33
S23	2141472.42	1190678.60

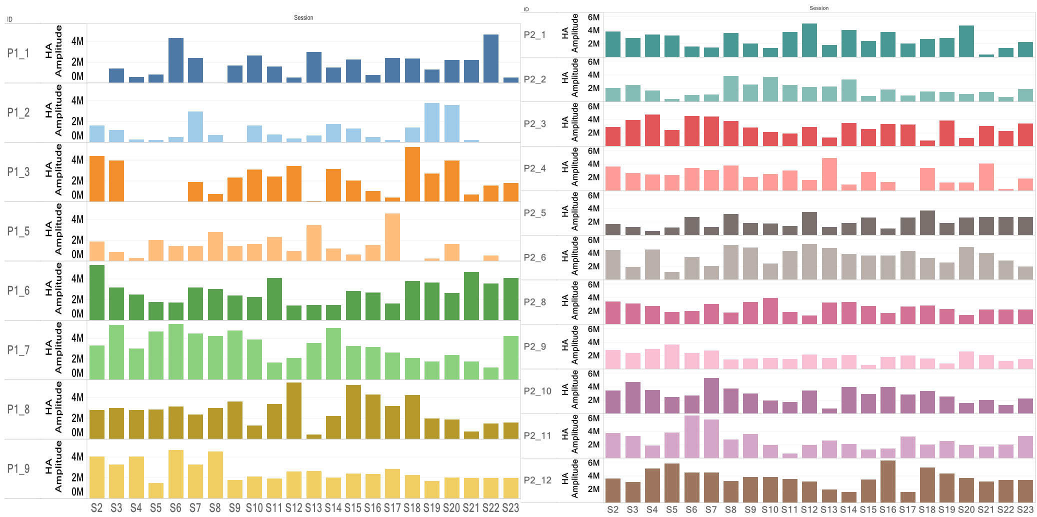


Figure 6.12. **Individual plot of mean high alpha amplitude vs. sessions for pinwheel game (*N* = 19).**

Assumption and Analysis

The Shapiro-Wilk test revealed that the mean high alpha amplitude per session was normally distributed, all *ps* > *0.05 (*except for S10*, p = 0.03).* The Mauchly's test did not provide any results as the sample size (*N* = 19) was less than repeated measurements (22 sessions). The Levene's test revealed that the homogeneity of variance was not violated for all sessions (all *ps* > 0.05) except for S2 and S23 (both *ps* = 0.04). The 2 × 22 Mixed ANOVA with Greenhouse–Geisser correction was done with groups and session as IV and the mean high alpha amplitude per session as DV.

Effects of Sessions and Groups on Mean High Alpha Amplitude

There was no significant interaction between groups and session on the mean high alpha amplitude across sessions, *N* = 19, *F* (8.09, 137.67) = 0.53, *p* = 0.82, η^2 = 0.03. There was no significant change in the mean high alpha amplitude across sessions for the pinwheel game, *N* = 19, *F* (8.09, 137.67) = 1.49, *p* = 0.16, η^2 = 0.08. There was no significant difference between G1 (*N* = 8) and G2 (*N* = 11) in the mean high alpha amplitude, *F* (1,17) = 1.61, *p* = 0.27, η^2 = 0.08.

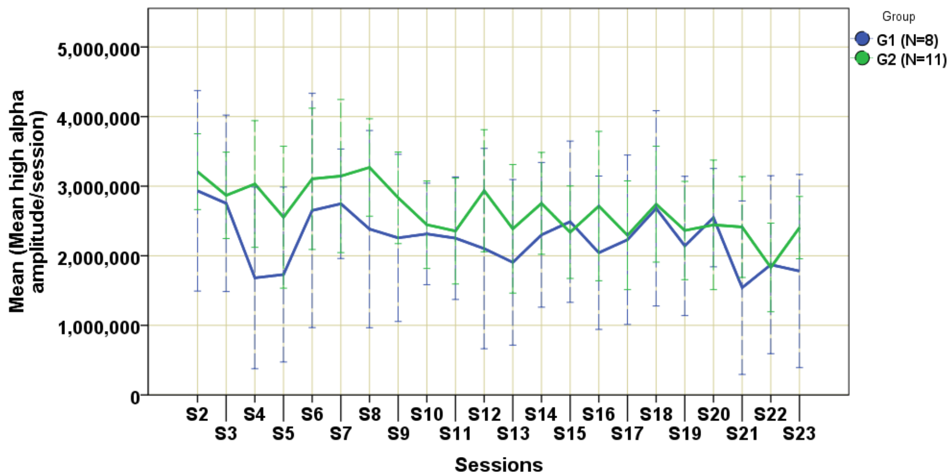


Figure 6.13. Mean (mean high alpha amplitude per session) vs. 22 sessions for G1 (*N*= 8) and G2 (*N* = 11) for pinwheel game.

6.4.4. Paraglider Game: Low Alpha

For research question (RQ 2.3.R.1), the mean low alpha amplitude per session was the DV. Sessions (S2 to S23) and groups (G1 and G2) were the IVs. The result reported in this section was for the paraglider game.

Missing Data and Extreme Outliers

For the paraglider game, 37 data points were missing, and two were eliminated as extreme outliers. Therefore, 9% of the data was missing from the entire dataset.

Descriptive Statistics

In this section, I present the descriptive statistics of the mean low alpha amplitude (no units) per session for the paraglider game. Tables 6.9 (below) provides the mean and standard deviation of the mean low alpha amplitudes for 19 participants from sessions S2 to S23. Figure 6.14 shows the data trend for each participant.

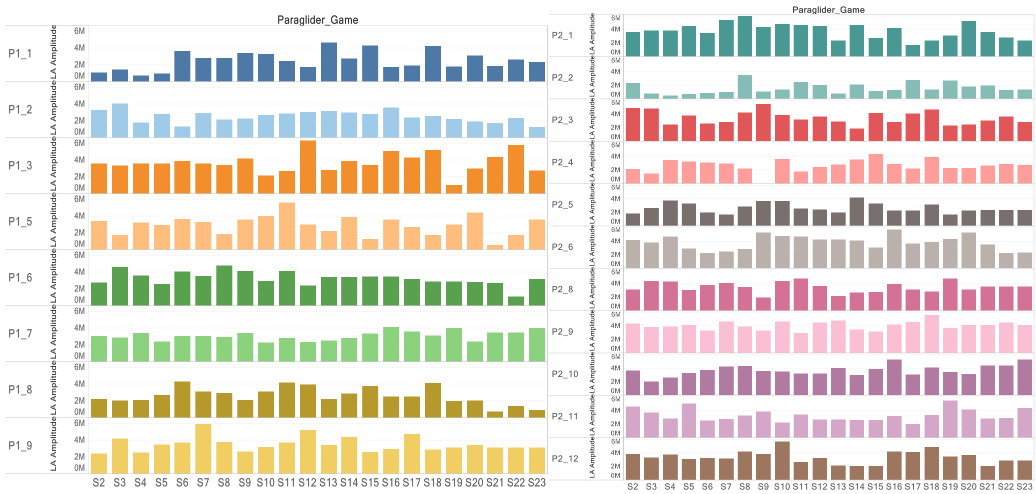


Figure 6.14. Individual plot of low alpha amplitude vs. sessions (S2 to S23) for paraglider game (N = 19).

Table 6.9. Descriptive statistics of (mean low alpha amplitude per session) for paraglider game (N = 19).

Sessions	Mean	Std. Deviation
S2	3164796.05	1012056.40
S3	3089359.12	1215948.86
S4	3006836.60	1123032.45
S5	3081739.66	1028112.70
S6	3077792.46	923857.29
S7	3339170.82	1168924.18
S8	3446460.30	1000959.78
S9	3269692.42	1347120.59
S10	3479475.37	1087791.67
S11	3407213.39	1012925.56
S12	3392677.80	1178170.71
S13	2900315.38	1008284.50
S14	3220608.03	821268.33
S15	3056132.53	911827.03
S16	3532414.37	1161328.83
S17	3088789.00	937155.64
S18	3508068.53	1141928.21
S19	3011105.58	1115856.84
S20	3184718.87	1042297.98

Sessions	Mean	Std. Deviation
S21	2762355.90	1099510.57
S22	2886971.14	1184203.55
S23	2920386.72	1090088.34

Assumption and Analysis

The Shapiro Wilk test revealed that the mean low alpha amplitude was normally distributed from S2 to S23 for 19 participants, all *p*s > 0.05. The Mauchly's test was undefined as the sample size was less than the repeated measurements. The Levene's test revealed that the homogeneity of variance had not been violated for all sessions (S2 to S23), all *p*s > 0.05. The 2 × 22 Mixed ANOVA with Greenhouse–Geisser correction was done with the groups and session as IV, and the mean low alpha potential per session as DV for the paraglider game.

Effects of Sessions and Groups on Mean Low Alpha Amplitude

There was no significant interaction between sessions and groups on the mean low alpha amplitude across the sessions, $N = 19$, $F\ (9.13,\ 155.35) = 1.19$, $p = 0.3$, $\eta^2 = 0.06$. There was no significant change in the mean low alpha amplitude across sessions, $N = 19$, $F\ (9.13,\ 155.35) = 1.19$, $p = 0.3$, $\eta^2 = 0.06$. There was no significant difference between G1 ($N = 8$) and G2 ($N = 11$) in the mean low alpha amplitude across sessions, $F\ (1,17) = 0.71$, $p = 0.4$, $\eta^2 = 0.04$.

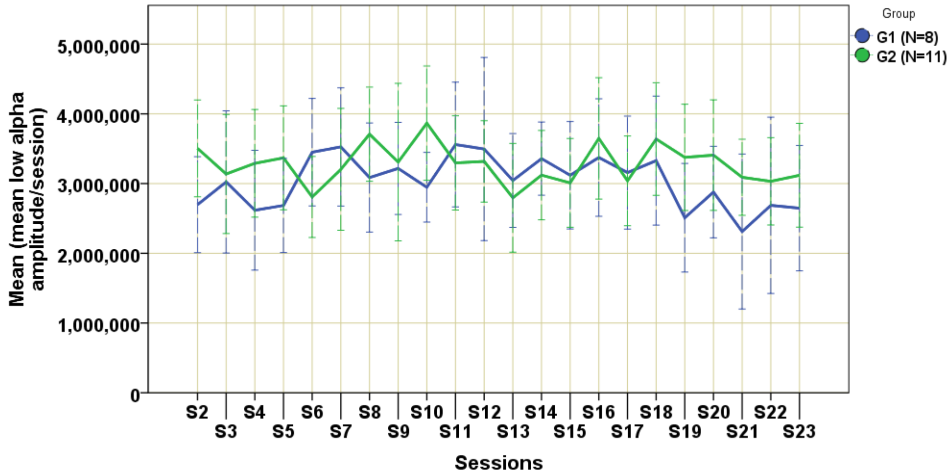


Figure 6.15. Mean (mean low alpha amplitude per session) vs. 22 sessions for G1 (*N*= 8) and G2 (*N* = 11) for paraglider game.

6.4.5. Paraglider Game: High Alpha

For research question (RQ 2.3.R.1), the mean high alpha amplitude (no units) per session was the DV. Sessions (S2 to S23) and groups (G1 and G2) were the IVs. The result reported in this section was for the paraglider game.

Missing Data and Extreme Outliers

For the paraglider game, 37 data points were missing and five data points were eliminated as extreme outliers. Therefore, 10% of the data was missing from the entire dataset.

Descriptive Statistics

In this section, I present the descriptive statistics of the mean high alpha amplitudes (no units) per session for the paraglider game. Tables 6.10 (below) provides the mean and standard deviation of the mean high alpha amplitudes for 19 participants from sessions S2 to S23. Figure 6.16 shows the data trend for each participant.

Table 6.10. Descriptive statistics of (mean high alpha amplitude per session) for paraglider game ($N = 19$).

Sessions	Mean	Std. Deviation
S2	2649539.92	1316560.37
S3	2630296.87	1102524.42
S4	1983197.90	1076480.10
S5	2632959.80	1015222.27
S6	2422759.89	1286549.10
S7	2650470.63	1437101.19
S8	2830613.61	1109456.91
S9	2286294.42	1316784.83
S10	2629165.00	1062752.19
S11	2219731.29	880336.91
S12	2106304.20	1149739.90
S13	2088645.78	945887.54
S14	2115121.97	1226626.45
S15	1924652.53	945093.66
S16	2402572.26	1168827.68
S17	2564939.74	1160285.93
S18	2531537.11	1326717.20
S19	2086670.11	1049579.09
S20	1952461.34	1024862.07
S21	1986271.78	1124148.46
S22	2120860.12	1172707.05
S23	1949480.46	1041270.31

Figure 6.16 Individual plot of high alpha amplitude vs. sessions (S2 to S23) for paraglider game (*N* = 19).

Assumptions and Analysis

The Shapiro Wilk test revealed that the mean high alpha amplitude was normally distributed from S2 to S23 for 19 participants, $p > 0.05$ (except for S20, $p = 0.04$). The Mauchly's test did not give *p*-value as the sample size was less than the repeated measurement. The Levene's test revealed that the homogeneity of variance test had not been violated for all sessions (all *ps* > 0.05), except for S9 ($p = 0.04$) and S12 ($p = 0.003$). The 2 × 22 Mixed ANOVA with Greenhouse–Geisser correction was done with groups and sessions as IV, and the mean high alpha amplitude per session as DV.

Effects of Sessions and Groups on Mean High Alpha Amplitude

There was no significant interaction between groups and session on the mean high alpha amplitude across sessions, $N = 19$, F (8.15, 138.66) = 1.26, $p = 0.26$, $\eta^2 = 0.06$. There was no significant change in mean high alpha amplitude across sessions for the paraglider game, $N = 19$, F (8.15, 138.66) = 1.79, $p = 0.08$, $\eta^2 = 0.09$. There was no significant difference between G1 (N = 8) and G2 (N = 11) in mean high alpha amplitude across sessions, F (1,17) = 0.78, p = 0.41, η^2 = 0.04.

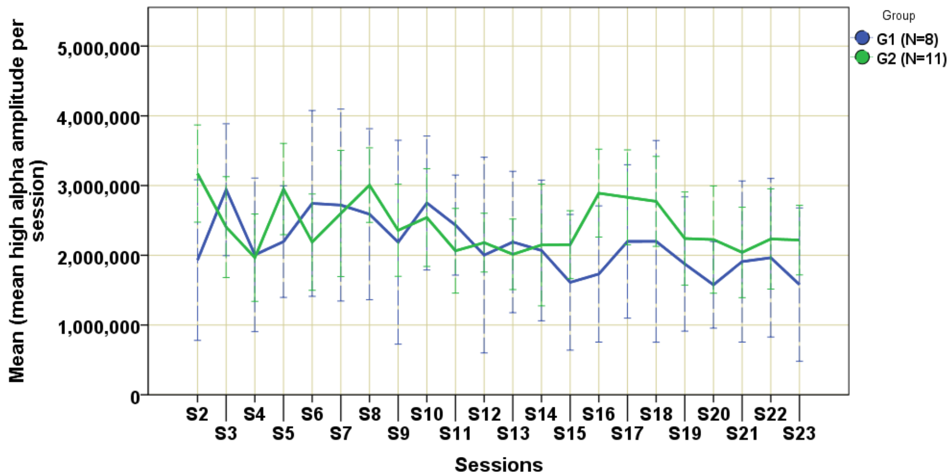


Figure 6.17. Mean (mean high alpha amplitude per session) vs. 22 sessions for G1 (*N*= 8) and G2 (*N* = 11) for paraglider game.

6.4.6. Stones Game: Low Beta

For RQ 2.3.A.1, mean low beta amplitude was the dependent measure. Sessions (S2 to S23) and groups (G1 and G2) were IVs. The analysis was conducted separately for the stones game alone.

Missing Data and Extreme Outliers

For the stones game, 42 data points were missing, and six data points were eliminated as extreme outliers. Therefore, 11% of the data was missing from the entire dataset.

Descriptive Statistics

In this section, I present the descriptive statistics of the mean low beta amplitude (no units) per session for the stones game. Tables 6.11 (below) provides the mean and standard deviation of the mean low beta amplitudes for 19 participants from sessions S2 to S23. Figure 6.18 shows the data trend for each participant.

Table 6.11. Descriptive statistics of (mean low beta amplitude per session) for stones game (N = 19).

Sessions	Mean	Std. Deviation
S2	1609712.21	1087934.37
S3	1640241.74	916767.31
S4	1327434.45	973019.47
S5	1475223.37	1036285.21
S6	1606755.47	848428.03
S7	1382014.47	961356.19
S8	1815877.53	1134025.56
S9	1365228.89	1055785.95
S10	1448922.68	794531.12
S11	1609686.00	877377.26
S12	1280891.16	709265.27
S13	1414288.84	798095.94
S14	1250320.21	598028.39
S15	1311191.47	722406.93
S16	1418839.79	951373.51
S17	1356104.74	1019103.15
S18	1304818.84	776539.84
S19	1170494.05	884776.64
S20	1404050.84	875176.47
S21	1037541.58	764273.63
S22	1233542.47	1120508.87
S23	1150833.53	895319.29

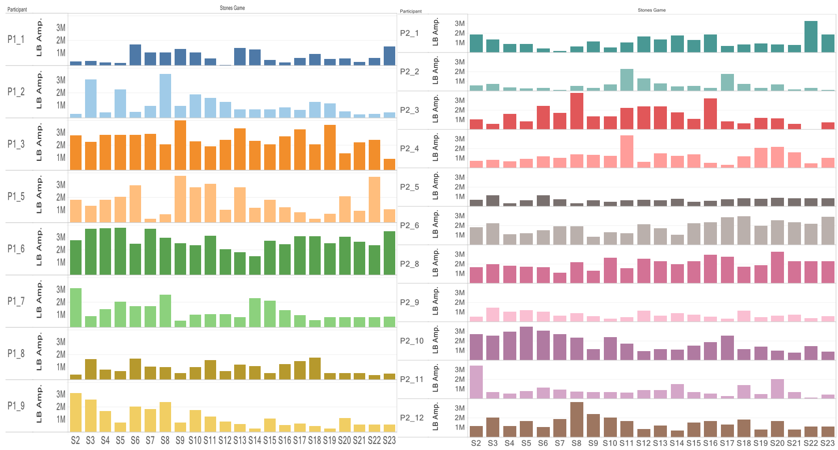


Figure 6.18. **Individual plot of low beta amplitude vs. sessions (S2 to S23) for stones game (*N* = 19).**

Assumptions and Analysis

The Shapiro Wilk test revealed that the mean low beta amplitude per session was normally distributed only for 12 out of 22 sessions with $p > 0.05$ (S3, S5, S6, S7, S8, S10, S11, S12, S14, S15, S16, and S18). The Levene's test revealed that the homogeneity of variance had not been violated between G1 and G2 for all the sessions, $p > 0.05$, except for S9 ($p = 0.01$). The Mauchly's test revealed that the assumption of sphericity for sessions was undefined as the number of subjects ($N = 19$) was less than the repeated measurement count ($P = 22$). The 2×22 Mixed ANOVA with Greenhouse–Geisser correction was done with groups and session as IV, and the mean low beta potential per session as DV for the stones game.

Effects of Sessions and Groups on Mean Low Beta Amplitude

There was no significant interaction between groups and session on the transformed mean low beta amplitude across sessions, $N = 19$, F (8.28, 140.79) = 0.83, $p = 0.57$, $\eta^2 = 0.04$. There was no significant change in the transformed mean low beta amplitude across sessions for the stones game, $N = 19$, F (8.28, 140.79) = 1.51, $p = 0.15$, $\eta^2 = 0.08$. There was no significant difference between G1 ($N = 8$) and G2 ($N = 11$) in the transformed mean low beta amplitude, F (1,17) = 0.8, $p = 0.38$, $\eta^2 = 0.04$.

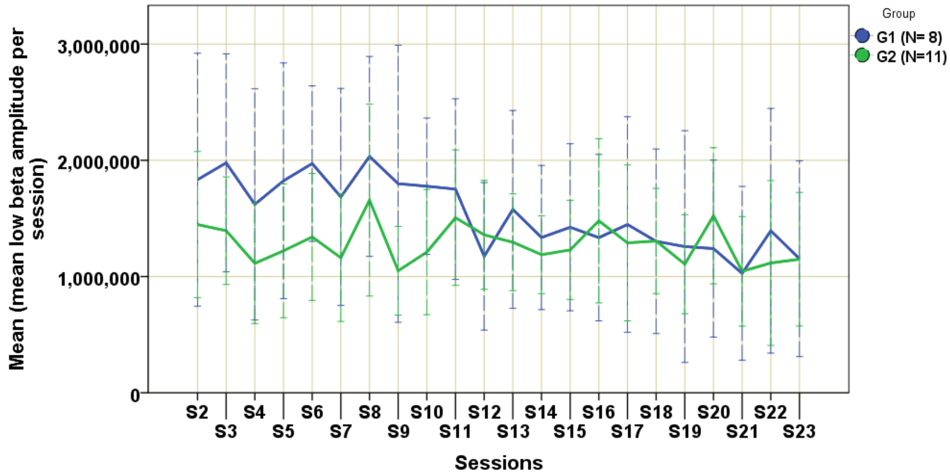


Figure 6.19. Mean (mean low beta amplitude per session) vs. 22 sessions for G1 (*N*= 8) and G2 (*N* = 11) for stones game.

6.4.7. Stones Game: Theta

For RQ 2.3.A.2, mean theta amplitude was the dependent measure. Sessions (S2 to S23) and groups (G1 and G2) were IVs. Sessions (S2 to S23) and groups (G1 and G2) were IVs. The analysis was conducted separately for the stones game alone

Missing Data and Extreme Outliers

For the stones game, 42 data points were missing, and seven data points were eliminated as extreme outliers. Therefore, 11% of the data was missing from the entire dataset.

Descriptive Statistics

In this section, I present the descriptive statistics of the mean theta amplitude (no units) per session for the stones game. Tables 6.12 (below) provides the mean and standard deviation of the mean theta amplitudes for 19 participants from sessions S2 to S23. Figure 6.20 shows the data trend for each participant.

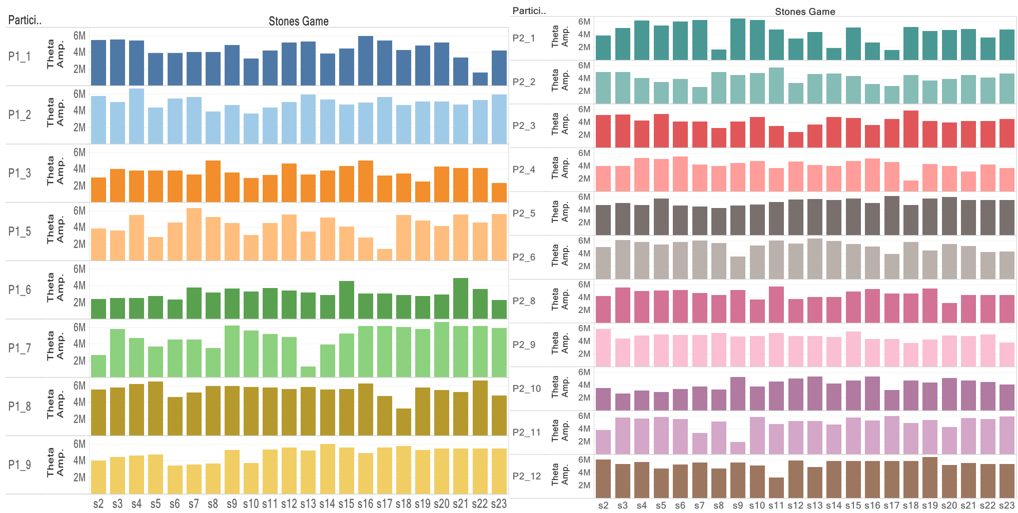


Figure 6.20. **Individual plot of theta amplitude vs. sessions (S2 to S23) for stones game.**

Table 6.12. **Descriptive statistics of (mean theta amplitude per session) for stones game (*N* = 19).**

Sessions	Mean	Std. Deviation
S2	4410004.42	1102994.14
S3	4767088.58	1032704.99
S4	4927937.05	1061421.34
S5	4553948.37	1089367.26
S6	4586060.11	958808.72
S7	4542917.16	1068067.58
S8	4241403.47	1071806.12
S9	4684765.26	1068649.89
S10	4482899.47	1035295.79
S11	4663326.95	880805.08
S12	4688763.84	974861.16
S13	4551850.05	1213297.57
S14	4553177.47	1088575.60
S15	5019262.32	561491.77
S16	4730319.74	1145149.24
S17	4332145.68	1466089.78
S18	4599490.84	1165115.28
S19	4698691.42	1009680.68
S20	4716643.90	945378.76

Sessions	Mean	Std. Deviation
S21	4823234.53	771205.84
S22	4616894.37	1113016.26
S23	4602319.58	1083299.49

Assumptions and Analysis

The Shapiro Wilk test revealed that the mean theta amplitude per session was normally distributed for each session for all 19 participants, $p > 0.05$, except for S12 and S16. The Mauchly's test revealed that the assumption of sphericity was undefined for within-subject analysis, as the number of children ($N = 19$) was less than the repeated measurement ($P = 22$). The Levine's test revealed that the homogeneity of variance had not been violated between groups for all the sessions ($p > 0.05$) except for S13 ($p = 0.01$) and S23 ($p = 0.02$). The 2 × 22 Mixed ANOVA with Greenhouse–Geisser correction was done with groups and session as IV and the mean theta amplitude per session as DV for the stones game.

Effects of Sessions and Groups on Mean Amplitude of Theta

There was no significant interaction between the groups and session on the mean theta amplitude across sessions, $N = 19$, F (8.17, 139.02) = 0.96, $p = 0.46$, $\eta^2 = 0.05$. There was no significant change in mean theta amplitude across sessions for the stones game, $N = 19$, F (8.17, 139.02) = 0.83, $p = 0.57$, $\eta^2 = 0.04$. There was no significant difference between groups G1 ($N = 8$) and G2 ($N = 11$) in the theta amplitude across sessions, F (1,17) = 0.26, $p = 0.61$, $\eta^2 = 0.01$.

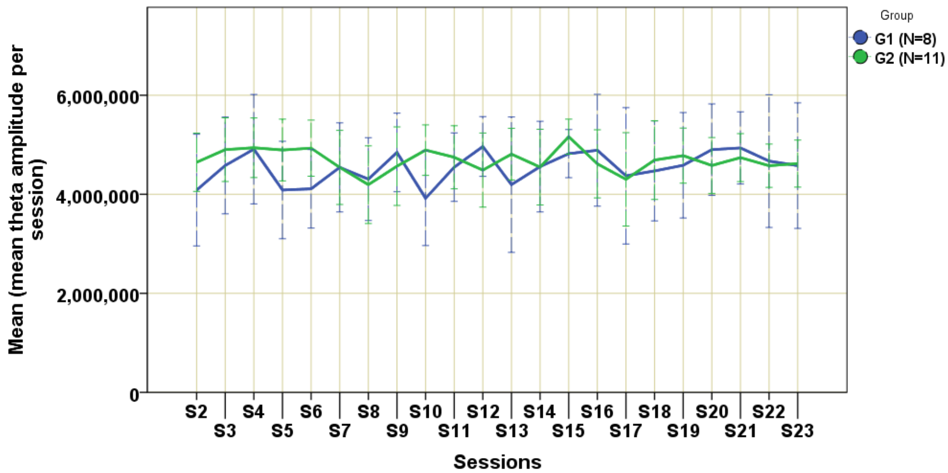


Figure 6.21. Mean (mean theta amplitude per session) vs. 22 sessions for G1 and G2 for stones game.

6.4.8. Summary

Unlike what I expected, both for the pinwheel and paraglider games, there was no improvement in the mean alpha amplitudes. This shows no solid improvement in the relaxation trend of the children. Similarly, for the stones game, there was no significant improvement in low beta amplitudes or decrease in theta amplitudes. This shows that there was no solid improvement in the attention trend of the children across sessions. As far as groups are concerned, as expected, the participant groups did not have any significant difference in their performance across the sessions.

6.5. Correlation between the Headset Relaxation & Attention Index Scores and EEG Amplitudes

(RQ 2.4.R.1) Is there a significant positive correlation between the mean relaxation score with the mean low alpha amplitude for each session in the pinwheel and paraglider games for all the participants?

(RQ 2.4.R.2) Is there a significant positive correlation between the mean relaxation score with the mean high alpha amplitude for each session in the pinwheel and paraglider games for all the participants?

(RQ 2.4.A.1) Is there a significant positive correlation between the mean relaxation score with the mean low beta amplitude for each session in the stones game for all the participants?

(RQ 2.4.A.2) Is there a significant negative correlation between the mean relaxation score with the mean theta amplitude for each session in the stones game for all the participants?

E: I expect to see a positive correlation between the dependent variable of relaxation measures used in RQ 2.2 and RQ 2.3 since an increase in alpha amplitudes signify an increase in relaxation for the pinwheel and paraglider games. I expect to see a correlation between the dependent variable of attention measures used in RQ 2.2 and RQ 2.3 as an increase in low beta amplitudes signifies an increase in attention, and a decrease in theta amplitude signifies an increase in attention.

6.5.1. Correlation of Relaxation Measurements for Pinwheel Game

In this section, I present the results of RQ 2.4.R.1 and RQ 2.4.R.2 for the pinwheel game.

Correlation between Mean Relaxation and Mean Low Alpha Amplitude

Outlier

Six outliers were removed.

Analysis

The Shapiro Wilk test revealed that the mean low alpha score was normally distributed ($p > 0.05$) and the mean relaxation score was not normally distributed ($p < 0.05$). The mean relaxation score did not show any linear or monotonic relationship with the between low alpha amplitudes. So, Kendall Tau-B was performed between mean relaxation and mean low alpha amplitude for the data points.

Results

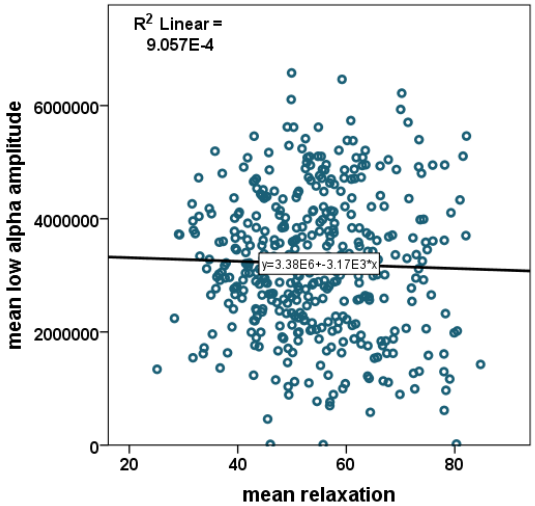


Figure 6.22. Scatterplot with mean low alpha amplitude and mean relaxation for 19 participants for pinwheel game.

There was a weak negative correlation between the mean relaxation score and mean low alpha potential, which was not significant (N = 413, τb = -0.014, p = 0.66)

Correlation between Mean Relaxation and Mean High Alpha Amplitude

Outlier

Six outliers were removed.

Analysis

The Shapiro Wilk test revealed that the mean high alpha score was normally distributed ($p > 0.05$) and the mean relaxation score was not normally distributed (both actual data and log transformation) ($p < 0.05$). The mean relaxation score did not show any linear or monotonic relationship with the high alpha amplitudes. So, Kendall Tau-B was performed between mean relaxation score and mean high alpha amplitude for the data points.

Results

There was a weak negative correlation between the mean relaxation and mean high alpha potential, which was statistically significant (N = 413, τb = -0.06, p = 0.03).

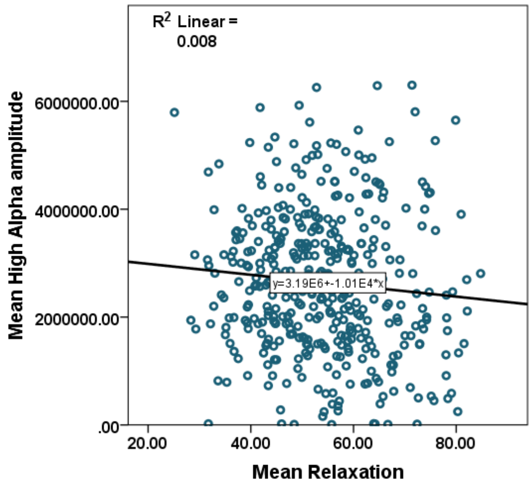


Figure 6.23. Mean high alpha amplitude vs. mean relaxation per session for 19 participants for pinwheel game.

6.5.2. Correlation of Relaxation Measurements for Paraglider Game

In this section, I present the results of (RQ 2.4.R.1) and (RQ 2.4.R.2).

Correlation between Mean Relaxation and Mean Low Alpha Amplitude

Outliers

Eight outliers were removed.

Analysis

The Shapiro Wilk test revealed that the mean low alpha amplitude was not normally distributed (p < 0.05). Similarly, the mean relaxation score was not normally distributed (p < 0.05). The mean relaxation score did not show any linear or monotonic relationship between low alpha amplitudes. So, Kendall Tau-B was performed between mean relaxation score and mean low alpha amplitude for the data points.

Results

There was a non-significant weak negative correlation between the mean relaxation score and mean low alpha amplitude (N = 372, τb = -0.04, p = 0.2).

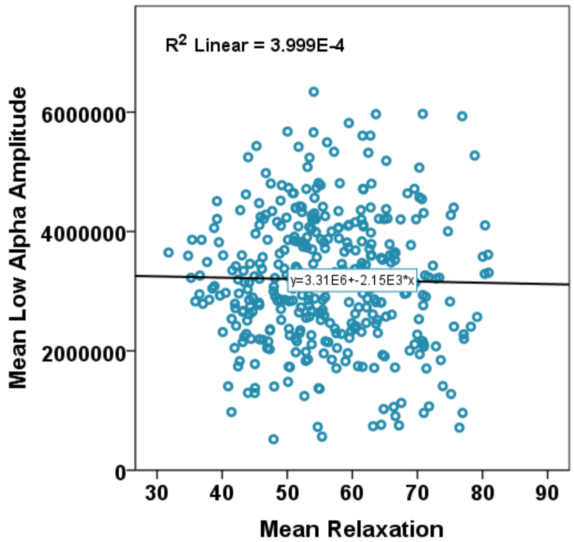


Figure 6.24. Scatterplot with mean low alpha amplitude and mean relaxation for 19 participants for paraglider game

Correlation between Mean Relaxation and Mean High Alpha Amplitude

Outliers

Eight outliers were removed.

Analysis

The Shapiro Wilk test revealed that the mean high alpha score was normally distributed ($p > 0.05$) and the mean relaxation score was not normally distributed (both actual data and log transformation) ($p < 0.05$). The mean relaxation score did not show any linear or monotonic relationship between high alpha amplitudes. So, Kendall Tau-B was performed between mean relaxation and mean high alpha amplitudes for the data points.

Results

There was a non-significant weak positive correlation between the mean relaxation score and mean high alpha amplitude ($N = 371$, $\tau b = 0.068$, $p = 0.19$).

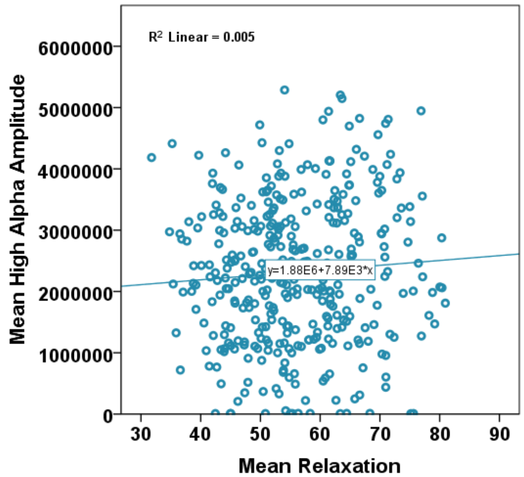


Figure 6.25. Scatterplot of mean high alpha amplitude and mean relaxation per session for 19 participants for paraglider game.

6.5.3. Correlation of Attention Measurements

In this section, I present the results of (RQ 2.4.A.1) and (RQ 2.4.A.2).

Correlation between Mean Attention and Mean Low Beta Amplitude

Outliers

Twelve outliers were removed.

Analysis

The Shapiro Wilk test revealed that the mean low beta amplitude was not normally distributed ($p < 0.05$) and the mean attention score was not normally distributed ($p < 0.05$). The mean attention score did not show any linear or monotonic relationship with the mean low beta amplitudes. Thus, Kendall Tau-B was performed between mean relaxation and mean amplitude for the data points.

Results

There was a non-significant negative correlation between the mean attention score and mean low beta amplitude (N = 363, τb = -0.045, p = 0.2).

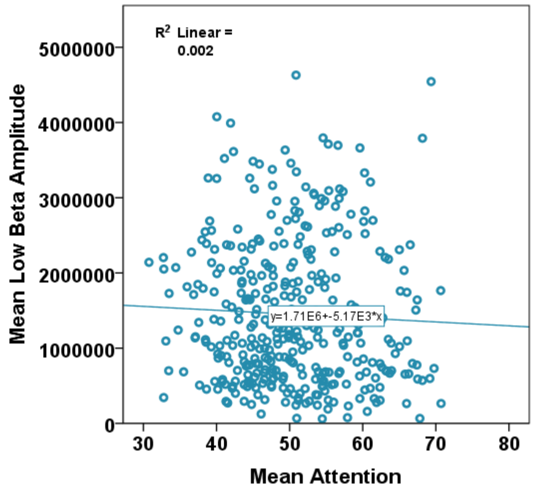


Figure 6.26. Scatterplot of mean low beta amplitude and mean attention per session for 19 participants for stones game.

Correlation between Mean Attention and Mean Theta Amplitude

Outliers

Seven outliers were removed.

Analysis

The Shapiro Wilk test revealed that the mean theta amplitude was not normally distributed ($p < 0.05$) and the mean attention score was not normally distributed ($p < 0.05$). The mean attention score did not show any linear or monotonic relationship with the mean theta amplitude. Therefore, Kendall Tau-B was performed between mean relaxation score and mean theta amplitude for the data points.

Results

There was a non-significant negative correlation between the mean relaxation score and mean theta amplitude ($N = 370$, $\tau b = -0.001$, $p = 0.91$).

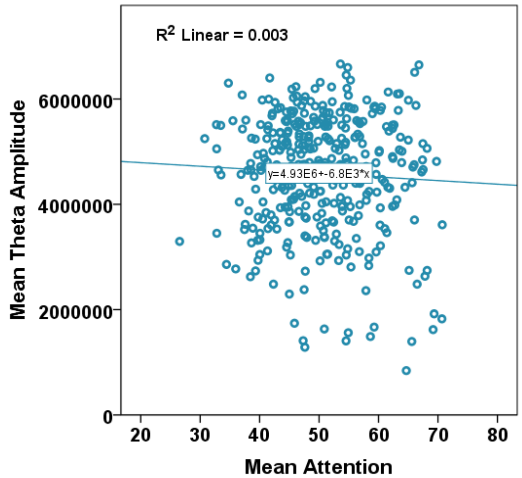


Figure 6.27. Scatterplot of mean theta amplitude and mean attention per session for 19 participants for stones game.

6.5.4. Summary

I did not get the expected results with respect to RQ 2.4. All correlation results were non-significant except RQ (2.4.R.2) for the pinwheel game. In the case of (RQ 2.4.R.2) for the pinwheel game, the mean relaxation per session had a significant weak negative correlation mean high alpha amplitude per session for the pinwheel game. However, I expected a positive correlation for this RQ as well. To conclude, there is no relationship between the dependent measures of relaxation and attention used in RQ 2.2 and RQ 2.3.

6.6. Token Collection Performance

(RQ 2.5) Did the time taken (in seconds) by the children to get five tokens in each session significantly reduce over 22 sessions (from session S2 to S23) for both groups of children (G1 and G2) for all three games (pinwheel, paraglider, stones) of the Mind-Full intervention?

Each token has two covariates, base_threshold and min_time. Assumption: There is a linear relationship between time taken to collect each token and their corresponding min_time.

***E*:** For all three games, the time taken to collect five tokens (TT) is expected to reduce over the intervention period for all the participants. It is expected that G1 and G2 do not have a difference in their ability to get tokens at a faster rate.

6.6.1. Pinwheel Game

For the pinwheel game, the dependent variable was the TT. The independent variables were sessions (S2 to S23), and groups (G1 and G2).

Missing Values and Extreme Outliers

There were 27 missing values with seven extreme outliers. Therefore, total data points excluded from the analysis amounted to 9% of the total data points produced.

Descriptive Statistics

In this section, I present the descriptive statistics of the TT per session for the pinwheel game. Tables 6.13 (below) provides the mean and standard deviation of the TT for 19 participants from sessions S2 to S23. Figure 6.28 shows the data trend for each participant.

Table 6.13. Descriptive statistics of TT per session for pinwheel game (*N* = 19)

Sessions	Mean	Std. Deviation
S2	38.81	13.09
S3	44.83	16.41
S4	36.23	11.67
S5	33.14	6.24
S6	31.76	8.65
S7	36.76	19.85
S8	36.56	10.45
S9	41.95	15.97
S10	40.19	12.09
S11	41.73	13.17
S12	37.67	12.24
S13	36.82	9.00
S14	36.47	8.36

Sessions	Mean	Std. Deviation
S15	42.80	14.12
S16	39.87	15.63
S17	36.33	10.26
S18	36.91	9.81
S19	39.27	11.65
S20	36.63	10.36
S21	33.81	9.00
S22	38.65	10.56
S23	35.53	10.50

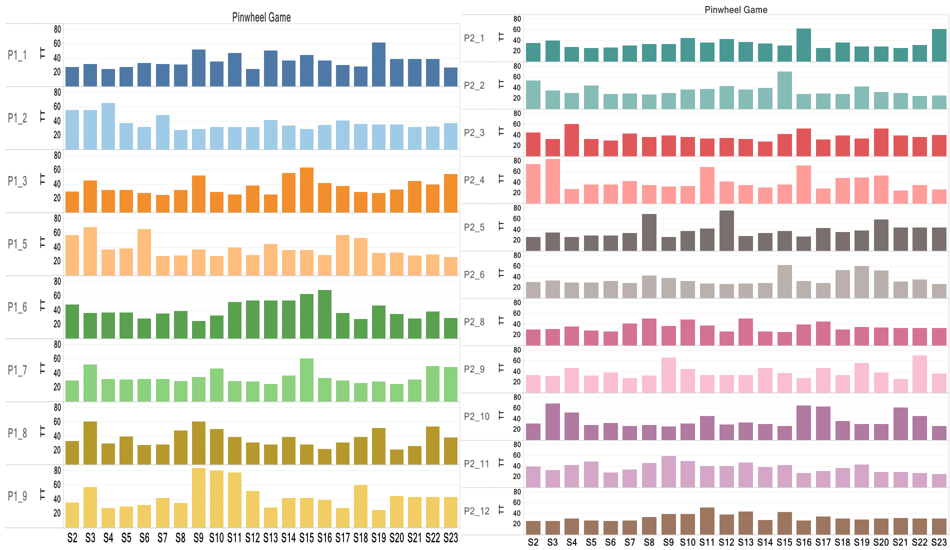


Figure 6.28. Individual plot of TT (in seconds) vs. sessions (S2 to S23) for pinwheel game (*N* = 19).

Assumptions and Analysis

The Shapiro Wilk test revealed that the TT was normally distributed for sessions S5, S13, S14, S19, and S20 ($p > 0.05$). All other sessions were not normally distributed. Therefore, a log transformation of the data was taken. Mauchly's test revealed that the sphericity of variance across sessions (within-subject) was undefined, as sample size ($N = 19$) was less than the repeated measurement counts ($N = 22$). Therefore, Greenhouse–Geisser correction was applied. The Levene's test revealed that the homogeneity of variance between groups G1 ($N = 8$) and G2 ($N = 11$) had not been violated, ($p > 0.05$). The 2 × 22 Mixed ANOVA with Greenhouse–Geisser correction was done with sessions and groups as IV and TT per session as DV.

Effects of Sessions and Groups on Time Taken to Collect Five Tokens in a Session.

There was no significant interaction between groups and sessions on the log transformed TT, $N = 19$, F (8.63, 146.72) = 0.79, $p = 0.61$, $\eta^2 = 0.04$. There was no significant change in log transformed TT across sessions for the pinwheel game, $N = 19$, F (8.63, 146.72) = 1.67, $p = 0.1$, $\eta^2 = 0.09$. There was no significant difference between the groups G1 ($N = 8$) and G2 ($N = 11$) in log transformed TT across sessions, F (1,17) = 0.88, $p = 0.36$, $\eta^2 = 0.04$.

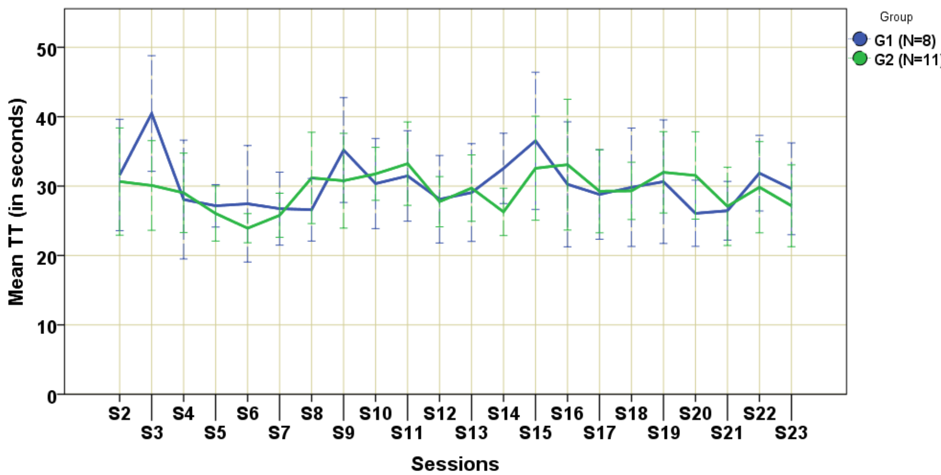


Figure 6.29. Mean TT (in seconds) vs. 22 sessions for G1 (*N* = 8) and G2 (*N* = 11) for pinwheel game. Error bar: 95% CI.

6.6.2. Paraglider Game

For the paraglider game, the dependent variable was the TT. The independent variables were sessions (S2 to S23), and groups (G1 and G2).

Missing Values and Extreme Outliers

There were 40 missing values with 27 extreme outliers. Therefore, total data points excluded from the analysis amounts to 16% of total data points collected.

Descriptive Statistics

In this section, I present the descriptive statistics of the TT per session for the paraglider game. Tables 6.14 (below) provides the mean and standard deviation of the TT for 19 participants from sessions S2 to S23. Figure 6.30 shows the data trend for each participant.

Table 6.14. Descriptive Statistics of TT per session for paraglider game (*N* = 19).

Sessions	Mean	Std. Deviation
S2	81.95	23.78
S3	87.18	25.22
S4	84.74	26.78
S5	82.49	30.48
S6	73.65	17.34
S7	72.30	13.68
S8	83.60	24.17
S9	79.12	27.34
S10	76.37	17.17
S11	87.85	28.64
S12	87.63	30.48
S13	85.81	27.68
S14	86.10	21.69
S15	80.55	19.84
S16	82.60	28.24
S17	79.07	22.51
S18	81.29	24.29
S19	83.04	30.82
S20	95.04	30.67
S21	80.62	17.41
S22	84.12	25.13
S23	75.36	19.03

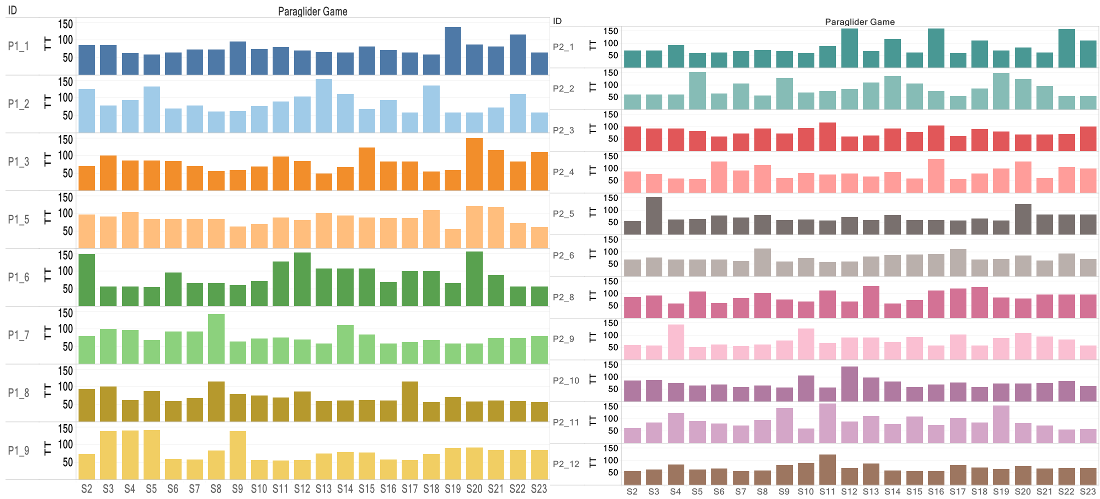


Figure 6.30. Individual plot of TT (in seconds) vs. sessions (S2 to S23) for paraglider game (*N* = 19).

Assumptions and Analysis

The Shapiro Wilk test revealed that the TT was normally distributed for S7, S8, S11, S13, S14, S15, and S19 ($p > 0.05$). All other sessions were not normally distributed. The Mauchly's test revealed that the sphericity of variance across sessions (within-subject) was undefined, as sample size ($N = 19$) was less than the repeated measurement counts ($N = 22$). Therefore, Greenhouse–Geisser correction was applied. The Levene's test revealed that the homogeneity of variance between G1 ($N = 8$) and G2 ($N = 11$) had not been violated, ($p > 0.05$), except for sessions S10, S16 and S20 ($p < 0.05$). The 2 × 22 Mixed ANOVA with Greenhouse–Geisser correction was done with sessions and groups as IV and TT per session as DV.

Effects of Session and Groups on Time Taken to Collect Five Tokens in a Session.

There was no significant interaction between groups and session on the TT, $N = 19$, F (8.45, 143.67) = 0.68, $p = 0.71$, $\eta^2 = 0.03$. There was no significant change in the TT across sessions for the paraglider game, $N = 19$, F (8.45, 143.67) = 0.86, $p = 0.55$, η^2 = 0. 04.. There was no significant difference between G1 ($N = 8$) and G2 ($N = 11$) in TT across sessions, F (1,17) = 0.12, $p = 0.72$, $\eta^2 = 0.007$.

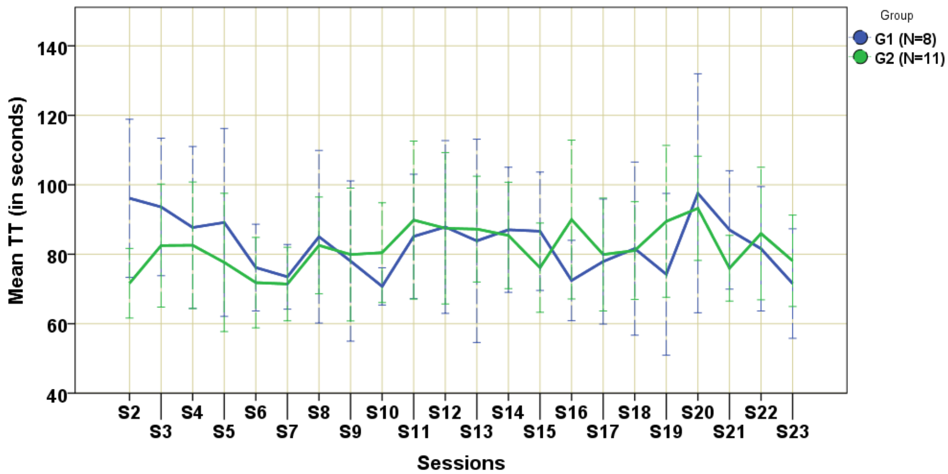


Figure 6.31. Mean TT (in seconds) vs. sessions (S2 to S23) for G1 (N = 8) and G2 (N = 11) for paraglider game. Error bar: 95% CI.

6.6.3. Stones Game

For the stones game, the dependent variable was the TT. The independent variables were sessions (S2 to S23), and groups (G1 and G2).

Missing Values and Extreme Outliers

There were 44 missing values with 20 extreme outliers. Therefore, total data points excluded in analysis amount to 15% of total data points collected.

Descriptive Statistics

In this section, I present the descriptive statistics of the TT per session for the stones game. Tables 6.13 (below) provides the mean and standard deviation of the TT for 19 participants from sessions S2 to S23. Figure 6.32 shows the data trend for each participant.

Table 6.15. Descriptive statistics of TT per session for stones game (*N* = 19).

Sessions	Mean	Std. Deviation
S2	453.16	94.19
S3	467.08	120.30
S4	522.89	96.15
S5	514.10	127.80
S6	544.34	133.64
S7	503.70	114.21
S8	469.81	128.63
S9	470.80	138.30
S10	500.81	142.34
S11	500.33	155.15
S12	510.23	113.23
S13	436.91	75.78
S14	507.09	171.80
S15	526.76	173.08
S16	496.63	172.36
S17	470.77	120.44
S18	544.55	162.02
S19	538.58	159.93
S20	525.61	108.02
S21	505.50	161.21
S22	550.86	155.08
S23	498.65	152.29

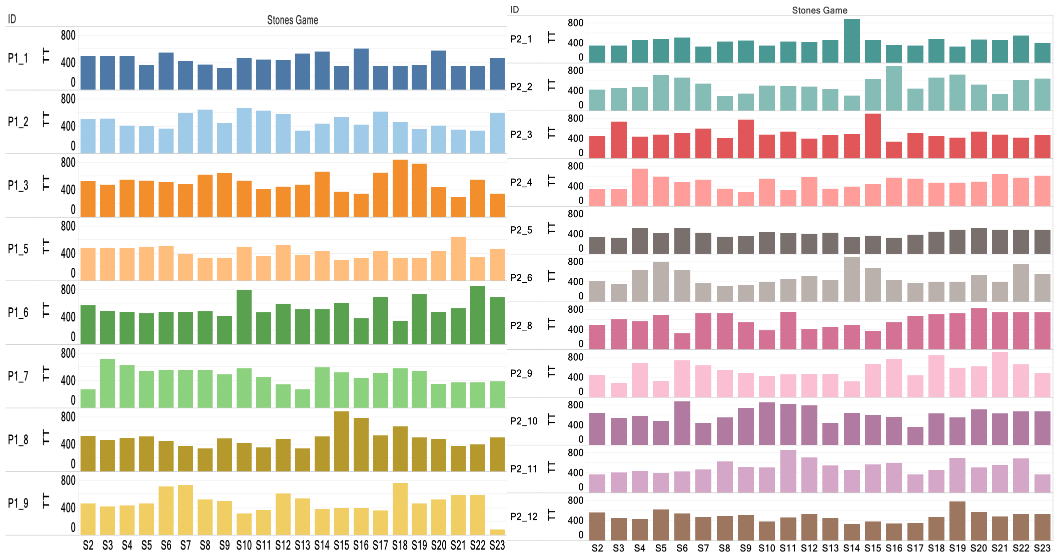


Figure 6.32. Individual plot of TT (in seconds) vs. sessions (S2 to S23) for stones game (*N*=19).

Assumptions and Analysis

The Shapiro Wilk test revealed that the TT was normally distributed for all sessions ($p > 0.05$) except for S10, S11, S14, and S17. The Mauchly's test revealed that the sphericity of variance across sessions (within-subject) was undefined, as sample size ($N = 19$) was less than the repeated measurement counts ($N = 22$). Therefore, Greenhouse–Geisser correction was applied. The Levene's test revealed that the homogeneity of variance between G1 ($N = 8$) and G2 ($N = 11$) had not been violated ($p >$ 0.05). The 2 × 22 Mixed ANOVA with Greenhouse–Geisser correction was done with sessions and groups as IV and TT per session as DV.

Effects of Session and Groups on Time Taken to Collect Five Tokens in a Session.

There was no significant interaction between groups and session on the TT, $N = 19$, F (9.29, 158.02) = 1.08, $p = 0.36$, $\eta^2 = 0.06$. There was no significant change in TT across sessions for the stones game, $N = 19$, F (9.29, 158.02) = 0.91, $p = 0.56$, $\eta^2 = 0.05$. There was no significant difference between G1 ($N = 8$) and G2 ($N = 11$) in TT, F (1,17) = 2.12, $p = 0.16$, $\eta^2 = 0.11$.

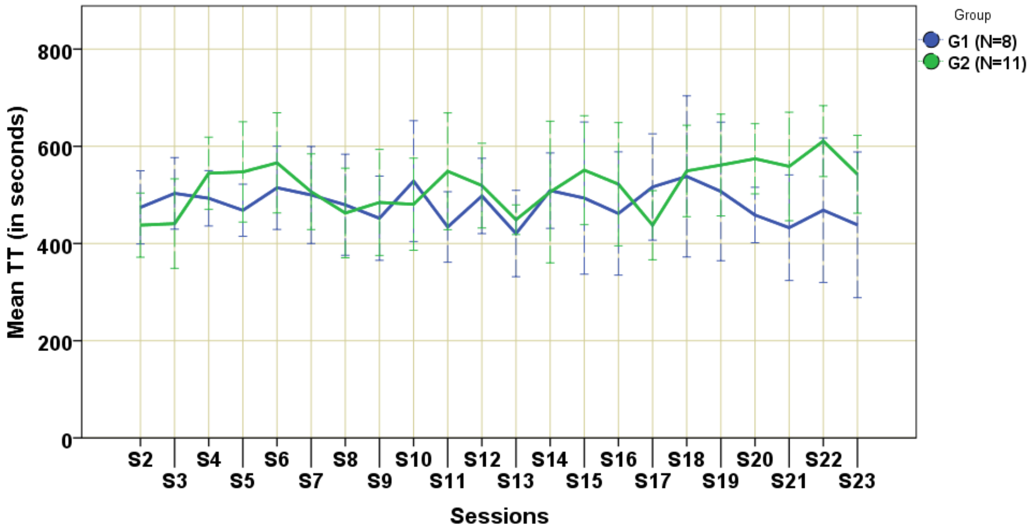


Figure 6.33. Mean TT (in seconds) vs. sessions (S2 to 23) for G1 (N = 8) and G2 (N = 11) for stones game. Error bar: 95% CI.

6.6.4. Summary

Contrary to my expectation, the TT or time taken to collect five tokens was not reduced across 22 sessions. However, as expected, there was no difference between groups in their change in time taken to collect five tokens across sessions.

Chapter 7. Discussion, Recommendations, and Conclusions

In this chapter, I discuss my data analysis methodology, what worked well and what did not, and alternate approaches based on previous research and the results. I then provide recommendations for future analysis of log data in other Mind-Full studies and similar NF studies.

7.1. What could the results mean?

I expected to see a gradual improvement in the children's relaxation and attention over 22 sessions. However, in the results, I did not observe any significant patterns. By observing the individual plots, I could see that the data had a lot of variability. There could be many factors producing this variability, including variability in the mental states of the children (especially when it was reported that they were experiencing trauma), headset accuracy, and other non-game related factors (Antle et al., 2015). The results I observed are quite different from the ones observed in some previous longitudinal NF research, where results showed significant improvement in attention by considering continuous sessions for analysis (Bakhshayesh et al., 2011 ; Egner et al., 2002 ; Hillard et al., 2013 ; Lubar et al., 1995). However, these researchers conducted their study with research-grade EEG headsets. My results are in alignment with other NF studies that had high fluctuations across sessions due to individual differences and personality (e.g. Dempster et al., 2009; Gruzelier 2014).

Despite my non-significant results, I could understand from the individual plots and my descriptive data (Section 6.1) that most of the children were able to play the game for more than the minimum session assigned i.e., 20 sessions. Only two children dropped out before 20 sessions. While looking at the individual plots of mean relaxation and attention (Figures 6.4, 6.5, and 6.6), I can observe that the mean relaxation and attention were mostly above the threshold of 40. According to Neurosky documentation, data above 40 represents a non-anxious/focused state. Looking at the individual plots of time taken to collect each token (Appendix A), I observe that the children were mostly able to complete the game by collecting five tokens per session, even in those cases where the difficulty level (min_time and base_threshold) increased with the sessions. This rare success and persistence suggests that the Mind-Full intervention was viable: regardless of the trauma experienced and the varying confounding factors such as low feeling days, these children could self-regulate to play the games and they were able to finish all their assigned sessions. Antle et al. (2015) have suggested that the pre-post analysis of the behavioural data (both qualitative and quantitative) showed improvement in the ability of the children to self-regulate after the intervention period. Therefore, there is a possibility that the children were able to self-regulate during gameplay and transferred their self-regulation skills from the Mind-Full game to their daily life. This explains why the children showed improvement in the behavioural data despite the high variability in the log data.

7.2. Data Pre-processing

7.2.1. Log Data Aggregation

The first and most important step of the analysis was to aggregate the log data collected in the Nepal study. The log data had two different files – the game file and the headset file. Both files had the session and participant information in their filename. The game file held information about the game events such as the type of game that started and ended and other events such as token collection and the token's parameters, min_time required and base_threshold. These events were asynchronous. The headset file held data provided by the headset at the rate of 1 Hz. It did not have any information about the game played. So, aggregation of these two files was necessary to identify which data belonged to which game. It was easy to aggregate both of these files for a participant and session combination by matching the filename, and the records in each file were aggregated using the timestamp of each record.

Recommendation 1 **(R1):** Once I consolidated all the data (for all sessions, for all users), it would be difficult to associate a record with its corresponding game, participant or session. Therefore, I recommend that the game type, participant ID, and session information should be logged for each record in the file. This would make it easier in the future to filter data based on a participant, session or game.

7.2.2. Noise Exclusion

It was very useful that the log data had Poor_Signal (noisy signal attribute) information. In this way, I could exclude the relaxation/attention values or EEG amplitudes with noisy signals while calculating mean values per session. The Nepal study's version of the Mind-Full software did not consider the headset attribute Poor_Signal to calculate the min_time required to hold relaxation above base_threshold. The headset mostly gave '0' for relaxation and attention when Poor_Signal > 0 (noisy signal instance), but it sometimes duplicated older values of relaxation and attention when the Poor_Signal > 0 (depicting noise) for a shorter duration (from two to 20 seconds based on manual observation). This did not give an accurate value for the time taken to collect a token. To exclude this confounding factor, I had to manually inspect and exclude those log data records that had both poor signal >0 and duplicate relaxation/attention scores.

(R2) For future studies, I would recommend that the Mind-Full application should consider the headset attribute Poor_Signal and exclude noisy data when calculating the min_time for achieving a token.

7.2.3. Reliability in Outlier and Missing Data Substitution

In this section, I describe how the rationale for substituting outlier and missing data. I followed a two-step procedure to exclude outliers. First, I used a standard box plot of each session to detect outliers for that session, followed by cross-verification of the outliers with the associated participant's trend to see if the participant had a similar range of data in other sessions or not. In this way, I ensured that the legitimate outliers (not due to instrumental/measurement error) were not excluded just because they were outside the box plot, as suggested by Osborne et al. (2004). I also ensured that I followed a reliable method to detect outliers before deciding to exclude them.

Similarly, to substitute missing values with the mean of nearest neighbours, I made sure that the differences in neighbouring values were not too high. For example, regarding the mean relaxation or attention score, I made sure that the difference in values was not more than 20 (per the guideline of the Neurosky eSense ranges[29]) and, for the time taken to collect tokens, that it was not twice the default times. In these cases, I substituted the mean of the four nearest neighbours. However, it was difficult to gauge the range for mean amplitudes as there were no proper amplitude ranges set by Neurosky. Therefore, I could not fully comprehend the range of the amplitudes and how they corresponded to the mental state of the children in the real-world, unlike how I related the relaxation score. This problem was more apparent when the amplitudes' scores were too high or low for a session, yet they were not outlying values. However, this problem can be eliminated in future studies by filtering and analyzing brainwave amplitudes from raw EEG signals. In doing so, there is no need to be dependent on the data calculated by the proprietary algorithm, and the researcher will have control over the frequency and amplitude ranges.

(R3) Firstly, before removing an outlier, investigate if the outlier is due to instrumental error or the performance of the child or some other unknown factor. Look for a trend in the data on both participant and session before excluding an outlier. Secondly, understand the data range before performing modifications to them. This will be helpful in deciding if a data point is an outlier or not. To substitute the missing data with nearest neighbour values, make sure that the difference in values between the neighbours does not have a high variability.

7.3. Dependent Measures Chosen

In this section, I discuss the DVs used in the current thesis, their distribution, and the alternate DVs that could be used for future studies, based on my analyses.

7.3.1. Noise in the Data

I expected noise to be random. The results also showed what I expected – the GQSP had no significant trends across sessions for all three games. This result of no adverse trend in the noise could possibly inform us that the headset was in a working condition throughout the training period. Another interesting aspect of noise pattern can

[29] http://developer.neurosky.com/docs/doku.php?id=esenses_tm

be observed between the games. From the descriptive statistics, both the pinwheel (mean GQSP range: 60–70) and stones (mean GQSP range: greater than 80) games had more noise than the paraglider game (mean GQSP range: greater than 80). Upon taking a closer look at the individual data sets, I observed that noise in the pinwheel games mostly occurred during the initial stages when establishing proper connectivity, but once a good signal was received, the pinwheel game had a noise pattern like that of the paraglider game. It is important to note that I removed the noise before analyzing the relaxation and attention measures. Since the NMM's proprietary algorithm calculates the noise parameter, there is obscurity about how these values are determined by the algorithm (i.e. Were muscular movements included to determine noise? Did they include eye blinks?); and so, I cannot say for certain that it entirely removed the influence of the noise from the data. This is a trade-off in the field study to choose the headset's portability and convenience to use over higher accuracy possible in the data with a research grade headset.

7.3.2. Relaxation and Attention Scores

The mean relaxation and mean attention scores of the consumer-grade EEG headsets as dependent measures has worked well with previous NF studies that had similar research goals to mine (Lee, 2009; Lim et al., 2012; Stinson & Arthur, 2013; Thomas et al., 2013). Based on these studies, I also considered the mean relaxation and attention measures per session as DVs. I calculated these values from the relaxation and attention scores given by the headset.

Even though I did not see any significant trend, it was interesting to see that the mean values ranged between 50–60 (SD between 4 and 14) for most of the sessions in all three games. Similar trends were also observed by Stinson et al. (2013), who stated that their mean attention indices were homogenous for all participants around the range of 50 and thus, could not show any significant difference. For my study, the mean relaxation and attention measures were above base_threshold 40. This could possibly mean that the children were able to use their bodies (breathing, relaxing, visual focusing) to self-regulate and be in a non-anxious and attentive state throughout the duration of the Mind-Full games.

(R4) Since most of the data points were above a certain range/threshold, it might be useful to analyze the time spent above a threshold instead of actual mean value. For example, if the children spend most of their session time in the range of 50 to 60 during the initial few sessions, it will be interesting to observe if the children were able to spend more time above higher threshold values (e.g., above 60) in the later sessions or towards the end of the training. Such trends might inform us that the children were able to spend more time at a higher threshold value as they keep practising the Mind-Full game.

7.3.3. Mean EEG Amplitudes

In section 2.2.3, it was seen that the increase in alpha amplitude implies an increase in relaxation. Similarly, an increase in low beta amplitude and decrease in theta amplitude implies an increase in attention. In section 2.4, it was seen that the mean amplitude per session can be treated as the dependent measure to observe change in relaxation/attention across sessions (Bakhshayesh et al., 2011 ; Egner et al., 2002 ; Hillard et al., 2013 ; Lubar et al., 1995). Most of these studies collected raw EEG signals (at the rate of 512 or 256 Hz), and filtered and recorded the amplitudes for the brainwaves before analyzing them. However, in my current study, the raw EEG signal was not collected. Therefore, I used the EEG amplitudes given by the NMM's proprietary algorithm.

There was no improvement in mean low alpha amplitude and mean high alpha amplitude across sessions. This contradicts previous research where an increasing trend in the mean alpha amplitude across continuous sessions was observed (Dempster et al. , 2009; Egner et al., 2002). Likewise, there was no significant decrease in theta amplitude and no significant increase in low beta amplitude to indicate improvement in attention. This, again, contradicts some previous research findings (Bakhshayesh et al., 2011; Hillard et al., 2013), wherein significant changes in amplitudes were observed by considering continuous sessions with children. However, we need to note that these studies were conducted using research grade EEG devices. Even though some of these studies shows improvement across sessions, when we analyze their DV trend across sessions, we can see fluctuations with the trends (Dempster et al. , 2009; Gruzelier, 2014). For those studies that have used NMM's headset EEG amplitudes, similar insignificant trends were observed for beta amplitudes (Robbins et al., 2014) and alpha amplitudes (Stinson et al., 2013).

Issue with Neurosky Calculated EEG Amplitudes

Even though I did not get the results that I expected, I would be overstating if I conclude that there was no improvement in relaxation and attention by playing the Mind-Full games. The non-significant result could also be due to high variability in the Neurosky calculated EEG amplitudes, which range between 10000-6M (no units). The proprietary algorithm of the Neurosky provided these amplitudes. Therefore, there is less transparency on how these values were calculated. Neurosky did not provide any proper documentation on expected range or units (Masasomeha, 2017).This made it difficult to make data preprocessing decisions such as excluding outliers or substituting missing data. Most other previous research used EEG amplitudes from the raw signals after excluding noise, eye blinks and other muscular movements (Bakhshayesh et al., 2011; Gevensleben et al., 2014; Hillard et al., 2013). For Mind-Full Nepal study, the eye blinks and the raw signals were not stored due to the limited storage capacity of the tablets and limited upload bandwidth in Nepal. So far, Stinson et al. (2013) had conducted their study using EEG amplitudes provided by Neurosky. However, they did not conduct their study over several sessions. Therefore, there is no evidence from previous research that could be used to show that the EEG amplitudes provided by the Neurosky were accurate and close to the EEG amplitudes derived from raw EEG signals for field conditions with children as participants.

(R5) If possible, for future analysis, it is recommended to collect, filter, and analyze raw signals (512 Hz) given by the Neurosky to see the change in amplitudes of different brainwave patterns. In analyzing EEG amplitudes, it is advised to record other parameters such as eye blinks (given by the Neurosky algorithm), and exclude data with high blink strength as noise.

Correlation between the Relaxation Score and Mean EEG Amplitudes

From RQ 2.5., it was seen that there was no positive correlation between mean relaxation scores and mean alpha amplitudes. There was no correlation between amplitudes of low beta or theta with mean attention. Further research is imperative to understand the correlation between Neurosky calculated EEG amplitudes and raw/actual EEG amplitudes for three main reasons: (1) the variation in data range is high; (2) there was no significant correlation between the dependent measures; and (3) there are no previous works that suggest the relationship between Neurosky calculated EEG amplitudes and raw EEG signal amplitudes. It is also necessary to conduct a study to understand if the Neurosky algorithm derives relaxation and attention from the alpha amplitude range, and low beta and theta amplitude range, respectively. I also suspect a possible internal validity threat (instrumental error) with respect to using Neurosky calculated EEG amplitudes unless the above-mentioned research question is answered.

(R6) To validate if the Neurosky algorithm calculates the relaxation/attention measure from the EEG amplitudes, I recommend conducting a study that involves correlational analysis between the relaxation/attention score and EEG amplitudes derived from the raw signal.

7.3.4. Time taken to Collect Tokens

After plotting the data, I found that almost 95% of the data had five tokens collected. Therefore, I set the time taken to collect five tokens as my DV. Even though I did not see any significant trend in the results, there were several confounding factors to this analysis.

First, the noisy data was not excluded by the Mind-Full application while computing the relaxation/attention measures. To handle this issue in the data analysis phase, I excluded the influence of the noise when it was more than six seconds. However, noisy data was still included in the token collection time when their influence was for less than six seconds. Therefore, it was difficult to get the accurate time taken to collect each token. However, it is important to note that it happened for less than 1% of the data, so it might not have a significant impact on my analysis.

Secondly, each token may have a different base_threshold and min_time, as these parameters could be dynamically changed by the teacher/counsellor using the calibration tablet. The counsellors did not change the base_threshold as much as they changed the min_time. Even though 99% of the tokens had their threshold set to the default value, the min_time parameter was changed both within a session and in-between sessions. As discussed in section 5.6.1, I normalised the time taken to collect each token. The min_time and the time taken to collect a token did not have a strong correlation. Thus, it was difficult to get accurate transformation about time taken to collect a token based on default min_time and base_threshold.

Thirdly, there were more outliers in the data for this dependent variable compared to other dependent variables. As seen in section 6.6, 15% of paraglider data and 16% of stones data were not included in this analysis (due to missing data or outliers). According to (Dong & Peng, 2013; Roth, 2002), missing data of more than 10% could have significant influence on analysis.

Lastly, there can be other factors such as loss of interest or stress when children were not able to collect token, which can deteriorate their performance. This can particularly be enhanced in the stones game, where the children need to collect five stones to get a token, which makes it five times as difficult to fill the jar compared to the previous games.

(R7) Consider alternate dependent variables that are not affected by the min_time and base_threshold.

(R8) In the Mind-Full software, add code to remove the influence of noisy signal on token collection. This could be done by stopping the timer when signal quality is poor or via other ways of excluding the time that signal quality is poor from the headset data.

7.3.5. Inconsistency in Missing Values

In certain cases, one of the dependent measures is missing even though the children played that session. For example, in session A, I would have a value for GQSP but would not have a value for mean relaxation per session due to connectivity issues. Similarly, I would have a value for mean relaxation per session but not have a value for the time taken to collect a token, as the children did manage to collect any tokens even though they played the game for five minutes. These were considered missing sessions for that RQ. This pattern might be interesting for future researchers, especially when there is a need to correlate different dependent measures such as mean relaxation and mean time taken to collect tokens. This correlation would help us understand if higher relaxation scores help children to collect tokens faster in the paraglider game.

7.4. Sampling Sessions

Revisiting the literature review, the session sampling was done in one of the four ways: (1) By considering (most of the) continuous sessions for analysis (Egner et al., 2002; Hillard et al., 2013; Lubar et al., 1995; McDonald, 1974; Takahashi et al., 2014); (2) by grouping few continuous sessions as blocks and use the block's mean value for analysis (Bakhshayesh et al., 2011); (3) by selecting equidistant singular sessions (e.g., S1, S5, S9, and S13) from the entire data set for analysis (Gevensleben et al., 2014); and (4) by comparing dependent values of the final few sessions with the initial few sessions for analysis (deBeus & Kaiser, 2011; Drechsler et al., 2007 ; Lim et al., 2012 ; Thomas et al., 2013).

As the first step of this analysis, I chose the first method, selecting most of the continuous sessions for analysis, and reserved the other methods such as analyzing data by grouping sessions into blocks or by comparing first and last few sessions for future work. I did not want to miss any significant effect of most of the sessions of the Mind-Full training on relaxation and attention measures and so I took this decision. Unlike the results of previous studies that had significant trends across sessions (Hillard et al., 2013; McDonald, 1974; Takahashi et al., 2014), I did not observe any significant improvement in my dependent measures. However, the previous studies collected data using research-grade EEG headset, while a consumer-grade EEG headset was used in the Mind-Full study. Stinson et al. (2014) did not observe significant improvement even in a single-session analysis by using the NMM headset. Sampling 22 sessions together introduced a lot of variability into the data. There was no increase or decrease in trend for any of the dependent variables across 22 sessions. It is highly likely that the children's mental states varied from day to day by contextual factors such as their daily activities, stressful situations at home before coming to school, behavioural challenges at school, and/or inconsistent counsellor coaching. Other factors that could likely have caused this variability in my results are data inaccuracy resulting from the field environment or children's head movement. These factors may have had some influence on the calm and attentive state of the children, which might have in turn affected their brainwave patterns and their performance across sessions. These factors were not recorded in the Mind-Full study data. There is a possibility that the high variability seen in my dataset may be due to these confounding factors. To confirm this, additional investigation is needed on the behaviour, activities, and movement of the children while participating in the Mind-Full training.

(R9) For future studies, it is better to log the daily behaviour and mental-state of the children, events that occurred before the start of the session, and other significant information such as the children's movement. This information will clarify the variability in the data and help researchers to make analysis decisions.

The two groups did not have any significant differences in their relaxation and attention trends across 22 sessions. However, when I see the trend plots, I observe that there might be a difference in performance between the groups for initial and final few sessions. For example, in all paraglider graphs, I see that G2 started better as compared to G1, but this trend collapsed with time because of the up-and-down trends of the consequent sessions. Since I took 22 measures for analysis, these details could have been insignificant due to high variability in the data. However, if I sample initial and final few sessions, it would be interesting to see if there is any significant difference in performance between the groups. As an initial step, sampling continuous sessions will help to understand the gradual trend caused by all the training sessions. For my data set, based on my observation of the data trends, I recommend method (4) for future work, where the first few and last few sessions are chosen for analysis. It will be helpful to understand if the children improved their self-regulation abilities at the end of the training, when compared to initial sessions. In method (3), sessions were grouped as blocks before considering them for analysis. This method would not be able to highlight the variability caused by the sessions and might not represent the contribution of all sessions. Method (2) (considering only equidistant singular sessions) would not work for my data as it relies on single sessions, which might have a lot of confounding factors.

(R10) After analyzing the patterns in continuous sessions, trends in initial few and final few sessions can be investigated to compare if there are any improvements. This method is followed in (deBeus & Kaiser, 2011; Drechsler et al., 2007 ; Lim et al., 2012 ; Thomas et al., 2013).

7.5. Inferential Analysis

I decided to use mixed ANOVA rather than other inferential methods such as regression because previous NF studies similar to the Mind-Full study design have used ANOVA for their analysis (Bakhshayesh et al., 2011 ; Egner et al., 2002 ; Gevensleben et al., 2014 ; Hillard et al., 2013 ; Kirenskaya et al., 2011 ; Lagopoulos et al., 2009 ; McDonald, 1974 ; Takahashi et al., 2014). Some previous works have used regression (Lee, 2009) or correlation (Lubar et al., 1995) to observe change in patterns across the training period. I did not report those results since there was no linear (or polynomial) trend observed in the individual data plots.

7.5.1. Sphericity Correction due to Sample Size

I can see that I had around 22 repeated measurements of dependent measures (S2 to S23). However, my sample size was 19. In Mauchly's test, the sphericity results were undefined and Greenhouse–Geisser correction was included while reporting the results as the sample size was less than the repeated measurement count. This could have lowered the statistical power of the analysis.

(R11) It is recommended, when investigating changes in dependent variables across the sessions, to use a higher sample size relative to repeated measurement counts.

7.5.2. Exclusion Criteria

ANOVA does not allow missing values, so I had to find a way to substitute missing data. As discussed in section 5.1, I had to exclude certain participants who had less than 15 sessions. I sampled sessions in such a way that most participants were included in the analysis. Therefore, I decided to sample up to 23 sessions even though a few participants had more than 23 sessions. These exclusion criteria were decided after observing the individual plots of many sessions for each participant. These issues could have been excluded in higher statistical models that accepts varying amounts of repeated measurements for each participant. For example, in regression, each participant can have different session counts. However, I did not choose these methods as I did not see any definite trend (e.g., linear or polynomial) in the individual plots of the participants. So, there was a trade-off between the inclusion of valid data and the valid statistical method that could capture most of the variability in the data.

7.5.3. Violation of Normality

I made sure that the Mixed ANOVA was used when the assumptions were met or when there was a slight deviation from the assumptions (e.g., five to six sessions were not normally distributed). If more than half of the session's data violated the assumptions (e.g., non-normal or heterogenous data), a non-parametric test was used. Running ANOVA with non-normal data might increase the chances of type-1 error (false positive) and overstate the significance. Since, we did not have any significant data, it should not have adverse effect on our results. Non-parametric test has less power than the parametric test[30]. For this reason, I chose to report the results with ANOVA instead of non-parametric test as they have less power. However, if I had seen a significant result, I would have cross checked the results with the equivalent non-parametric test.

7.5.4. Other Limitations

Firstly, the scope of the analysis was restricted to observing the data patterns between most of the sessions (N = 22). Even though a significant difference in such patterns would help me make stronger claims, it is highly difficult to capture patterns in such a case from a field environment using NF data with children. This data analysis was more conservative, by considering most of the sessions for repeated measurements, etc. However, it is important to do this analysis as a first step to understand the overall data

[30] http://blog.minitab.com/blog/understanding-statistics/data-not-normal-try-letting-it-be-with-a-nonparametric-hypothesis-test

pattern before trying to explore special cases in the data (e.g., trend in first few or last few sessions, within-session improvements, individual performance, etc.). Some of the previous studies have considered baseline measurements (data collected without NF) and have studied and observed improvements in the relaxation and attention trends within a session (Gruzelier et al., 2014). These analyses are kept for future work.

Secondly, the data analysis was performed with the assumption that the data from the NMM headset values would be accurate, even though there was no strong evidence from previous research regarding the accuracy of the headset data under field conditions. In such cases, using the same device to control the game and to validate the game might be a limitation as the errors in the headset data might have an influence on both game control and MF success evaluation.

(R12) Apart from analyzing the same data that controls the game to provide NF training, collect and analyze additional objective measures (such as heart rate variability through wrist watches to detect anxiety, standard ADHD and EF tests, etc.) to evaluate the success of the NF training. This method might exclude the bias of the headset data in the evaluation process.

Finally, the data pre-processing, session sampling and the statistical procedures were mostly adopted from the previous NF studies that were discussed in the literature review. There might be other alternate approaches in the statistical domain to handle this data. I leave this approach for future exploration.

7.6. Conclusion

In this work, I present a novel method of performing a review based data analysis approach for analyzing the data from a Mind-Full like study, i.e. data from a consumer-grade headset from a multi-sessional field study. I analyzed the log data from the field-study that was conducted in Nepal over from November to March 2014-2015. The goal was to help children living in extreme poverty to self-regulate their calmness and attention. Based on previous research and my own rationale, I (1) developed dependent measures for relaxation, attention, and game performance from the log data, (2) determined how to sample sessions to see if there were significant changes or patterns in the dependent variables across continuous sessions, (3) developed an approach to handle data constraints such as noisy data, missing sessions, and outliers , and (4) applied appropriate data analysis methods to identify significant trends in the dependent measures across sessions.

To estimate the noise present in the data, I considered the percentage of time that had a good signal as the dependent measure. I considered mean relaxation per session, mean low alpha amplitude per session, and mean high alpha amplitude per session as the measure of relaxation. For attention, mean attention per session, mean low beta per session, and mean theta per session were considered as dependent variables. For game performance, the time taken to collect five tokens was considered as the dependent variable for each game, after normalizing the time taken to collect each token based on the set difficulty parameters. For each participant, I excluded extreme outliers after comparing the outlying data point with the data points of other sessions. For missing data and outliers, I substituted them with the mean score of the relevant neighbouring points. I considered 22 continuous sessions (sessions 2 to 23) for analysis as most participants had data for them. Relaxation measures were separately analyzed for the pinwheel and paraglider games, and attention measures were analyzed for the stones game, as each game had different goals.

Based on my analysis, I found that the noise occurrence was random and all noisy data was excluded from further analysis. I found that there was no significant trend observed for relaxation and attention in any of the dependent variables used for all three games while considering 22 continuous sessions for analysis. There was a lot of variability in the data. These variabilities could be possibly due to other confounding factors such as children's daily mental and emotional states, how they were treated at home before coming to school, headset data quality and accuracy with children as participants under field environment, and so on. Despite the variability, the children were able to self-regulate to play the game and complete the set goals of the NF training, which shows that the game-play was viable. There was no significant correlation between different measures of relaxation. Similarly, the different measures of attention did not correlate as well. This calls for validation of the accuracy of headset-calculated EEG amplitudes when compared to EEG amplitudes derived from raw signals in a field environment set-up with children. For future NF studies, I recommended to derive the dependent measures from the raw signals provided by the headset for analysis. Despite taking a research review based approach to data analysis, we found no significant patterns. It might be too soon to get results with this type of consumer headset for our experimental conditions. I recommend continuing to rely on other subjective and objective behavioral measures to evaluate NF interventions for children under field conditions.

References

Anderson, P. (2002). Assessment and Development of Executive Function (EF) during Childhood. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *8*(2), 71–82. https://doi.org/10.1076/chin.8.2.71.8724

Antle, A. N., Chesick, L., Levisohn, A., Sridharan, S. K., & Tan, P. (2015). Using Neurofeedback to Teach Self-regulation to Children Living in Poverty. In Proceedings of the 14th International Conference on Interaction Design and Children (pp. 119–128). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2771852

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using Electroencephalography to Measure Cognitive Load. *Educational Psychology Review*, *22*(4), 425–438.

Armsworth, M. W., & Holaday, M. (1993). The Effects of Psychological Trauma on Children and Adolescents. *Journal of Counseling & Development*, *72*(1), 49–56.

Baars, B. J., & Gage, N. M. (2010). Chapter 8 - Consciousness and attention. In *Cognition, Brain, and Consciousness (Second Edition)* (pp. 238–303). London: Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780123750709000085

Bakhshayesh, A. R., Hänsch, S., Wyschkon, A., Rezai, M. J., & Esser, G. (2011). Neurofeedback in ADHD: a single-blind randomized controlled trial. *European Child & Adolescent Psychiatry*, *20*(9), 481–491. https://doi.org/10.1007/s00787-011-0208-y

Bellis, M. D. D. (2001). Developmental traumatology: The psychobiological development of maltreated children and its implications for research, treatment, and policy. *Development and Psychopathology*, *13*(3), 539–564.

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive Functions after Age 5: Changes and Correlates. Developmental Review: DR, 29(3), 180–200.

Broughton, R., & Hasan, J. (1995). Quantitative topographic electroencephalographic mapping during drowsiness and sleep onset. Journal of Clinical Neurophysiology, 12(4), 372-hyhen.

Chabot, R. J., Merkin, H., Wood, L. M., Davenport, T. L., & Serfontein, G. (1996). Sensitivity and specificity of QEEG in children with attention deficit or specific developmental learning disorders. Clinical Electroencephalography, 27(1), 26-34.

Belkofer, C. M., & Konopka, L. M. (2008). Conducting art therapy research using quantitative EEG measures. Art Therapy, 25(2), 56-63.

Crowley, K., Sliney, A., Pitt, I., & Murphy, D. (2010). Evaluating a brain-computer interface to categorise human emotional response. In *2010 10th IEEE International Conference on Advanced Learning Technologies* (pp. 276–278). Ieee. Retrieved from http://www.computer.org/csdl/proceedings/icalt/2010/4055/00/4055a276-abs.html

deBeus, R. J., & Kaiser, D. A. (2011). Chapter 5 - Neurofeedback with Children with Attention Deficit Hyperactivity Disorder: A Randomized Double-Blind Placebo-Controlled Study. In *Neurofeedback and Neuromodulation Techniques and Applications* (pp. 127–152). San Diego: Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780123822352000056

Dempster, T., & Vernon, D. (2009). Identifying Indices of Learning for Alpha Neurofeedback Training. *Applied Psychophysiology and Biofeedback*, *34*(4), 309. https://doi.org/10.1007/s10484-009-9112-3

Diamond, A., & Lee, K. (2011). Interventions Shown to Aid Executive Function Development in Children 4 to 12 Years Old. *Science*, *333*(6045), 959–964. https://doi.org/10.1126/science.1204529

Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*. https://doi.org/10.1186/2193-1801-2-222

Drechsler, R., Straub, M., Doehnert, M., Heinrich, H., Steinhausen, H.-C., & Brandeis, D. (2007). Controlled evaluation of a neurofeedback training of slow cortical potentials in children with Attention Deficit/Hyperactivity Disorder (ADHD). *Behavioral and Brain Functions*, *3*(1), 35. https://doi.org/10.1186/1744-9081-3-35

Egner, T., Strawson, E., & Gruzelier, J. H. (2002). EEG Signature and Phenomenology of Alpha/theta Neurofeedback Training Versus Mock Feedback. *Applied Psychophysiology and Biofeedback*, *27*(4), 261. https://doi.org/10.1023/A:1021063416558

Empson, J. (1986). Human brainwaves: The psychological significance of the electroencephalogram. Springer.

Changyong, F. E. N. G., Hongyue, W. A. N. G., Naiji, L. U., Tian, C. H. E. N., Hua, H. E., & Ying, L. U. (2014). Log-transformation and its implications for data analysis. Shanghai archives of psychiatry, 26(2), 105.

Fuchs, T., Birbaumer, N., Lutzenberger, W., Gruzelier, J. H., & Kaiser, J. (2003). Neurofeedback treatment for attention-deficit/hyperactivity disorder in children: a comparison with methylphenidate. Applied psychophysiology and biofeedback, 28(1), 1-12.

Gevensleben, H., Kleemeyer, M., Rothenberger, L. G., Studer, P., Flaig-Röhr, A., Moll, G. H., … Heinrich, H. (2014). Neurofeedback in ADHD: Further Pieces of the Puzzle. *Brain Topography*, *27*(1), 20–32. https://doi.org/10.1007/s10548-013-0285-y

Goldstein, S., Naglieri, J. A., Princiotta, D., & Otero, T. M. (2014). Introduction: A History of Executive Functioning as a Theoretical and Clinical Construct. In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of Executive Functioning* (pp. 3–12). Springer New York. https://doi.org/10.1007/978-1-4614-8106-5_1

Government of Canada, I. A. P. on R. E. (2016, February 5). Interagency Advisory Panel on Research Ethics. Retrieved March 22, 2017, from http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/chapter2-chapitre2/#ch2_en_a2.5

Gruzelier, J., Egner, T., & Vernon, D. (2006). Validating the efficacy of neurofeedback for optimising performance. In C. N. and W. Klimesch (Ed.), *Progress in Brain Research* (Vol. 159, pp. 421–431). Elsevier. Retrieved from http://www.sciencedirect.com/science/article/pii/S0079612306590272

Gruzelier, J. H., (2014). EEG-neurofeedback for optimising performance. III: A review of methodological and theoretical considerations. Neuroscience & Biobehavioral Reviews, 44, 159–182. https://doi.org/10.1016/j.neubiorev.2014.03.015

Hillard, B., El-Baz, A. S., Sears, L., Tasman, A., & Sokhadze, E. M. (2013). Neurofeedback Training Aimed to Improve Focused Attention and Alertness in Children with ADHD A Study of Relative Power of EEG Rhythms Using Custom-Made Software Application. *Clinical EEG and Neuroscience*, *44*(3), 193–202. https://doi.org/10.1177/1550059412458262

Huang, J., Yu, C., Wang, Y., Zhao, Y., Liu, S., Mo, C., … Shi, Y. (2014). FOCUS: enhancing children's engagement in reading by using contextual BCI training sessions. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1905–1908). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2557339

Johnstone, S. J., Blackman, R., & Bruggemann, J. M. (2012). EEG From a Single-Channel Dry-Sensor Recording Device. *Clinical EEG and Neuroscience*, *43*(2), 112–120. https://doi.org/10.1177/1550059411435857

Judd, C. M., McClelland, G. H., & Culhane, S. E. (1995). Data analysis: Continuing issues in the everyday analysis of psychological data. *Annual Review of Psychology*, *46*(1), 433–465.

Keene, O. N., (1995). The log transformation is special. Statistics in Medicine, 14(8), 811–819. https://doi.org/10.1002/sim.4780140810

Kirenskaya, A. V., Novototsky-Vlasov, V. Y., & Zvonikov, V. M. (2011). Waking EEG Spectral Power and Coherence Differences Between High and Low Hypnotizable Subjects. *International Journal of Clinical and Experimental Hypnosis*, *59*(4), 441–453. https://doi.org/10.1080/00207144.2011.594744

Kropotov, J. D., Grin-Yatsenko, V. A., Ponomarev, V. A., Chutko, L. S., Yakovenko, E. A., & Nikishena, I. S. (2005). ERPs correlates of EEG relative beta training in ADHD children. *International Journal of Psychophysiology*, *55*(1), 23–34. https://doi.org/10.1016/j.ijpsycho.2004.05.011

Lagopoulos, J., Xu, J., Rasmussen, I., Vik, A., Malhi, G. S., Eliassen, C. F., … Ellingsen, (2009). Increased Theta and Alpha EEG Activity During Nondirective Meditation. The Journal of Alternative and Complementary Medicine, 15(11), 1187–1192. https://doi.org/10.1089/acm.2009.0113

Lane, D. M. (2017). Log Transformations. Retrieved March 22, 2017, from http://onlinestatbook.com/2/transformations/log.html

Lee, K. (2009). Evaluation of attention and relaxation levels of archers in shooting process using brain wave signal analysis algorithms. 감성과학, *12*(3), 341–350.

Legewie, H., Simonova, O., & Creutzfeldt, O. D. (1969). EEG changes during performance of various tasks under open- and closed-eyed conditions. *Electroencephalography and Clinical Neurophysiology*, *27*(5), 470–479. https://doi.org/10.1016/0013-4694(69)90187-4

Leins, U., Goth, G., Hinterberger, T., Klinger, C., Rumpf, N., & Strehl, U. (2007). Neurofeedback for Children with ADHD: A Comparison of SCP and Theta/Beta Protocols. *Applied Psychophysiology and Biofeedback*, *32*(2), 73–88. https://doi.org/10.1007/s10484-007-9031-0

Leite, J., Maria De Lourdes, V. S., Sartori, V. A., & Andreatini, R. (1999). The video-recorded Stroop Color-Word Test as a new model of experimentally-induced anxiety. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *23*(5), 809–822.

Liew, J., Lench, H. C., Kao, G., Yeh, Y.-C., & Kwok, O. (2014). Avoidance temperament and social-evaluative threat in college students' math performance: a mediation model of math and test anxiety. *Anxiety, Stress, & Coping*, *27*(6), 650–661.

Lim, C. G., Lee, T. S., Guan, C., Fung, D. S. S., Zhao, Y., Teng, S. S. W., … Krishnan, K. R. R. (2012). A Brain-Computer Interface Based Attention Training Program for Treating Attention Deficit Hyperactivity Disorder. *PLOS ONE*, *7*(10), e46692. https://doi.org/10.1371/journal.pone.0046692

Lim, C. G., Lee, T.-S., Guan, C., Sheng Fung, D. S., Cheung, Y. B., Teng, S. S. W., … Krishnan, K. R. (2010). Effectiveness of a brain-computer interface based programme for the treatment of ADHD: a pilot study. *Psychopharmacology Bulletin*, *43*(1), 73–82.

Liu, N.-H., Chiang, C.-Y., & Hsu, H.-M. (2013). Improving Driver Alertness through Music Selection Using a Mobile EEG to Detect Brainwaves. *Sensors (Basel, Switzerland)*, *13*(7), 8199–8221. https://doi.org/10.3390/s130708199

Lubar, J. F., Swartwood, M. O., Swartwood, J. N., & O'Donnell, P. H. (1995). Evaluation of the effectiveness of EEG neurofeedback training for ADHD in a clinical setting as measured by changes in T.O.V.A. scores, behavioral ratings, and WISC-R performance. *Biofeedback and Self-Regulation*, *20*(1), 83–99.

Lubar, J. O., & Lubar, D. J. F. (1984). Electroencephalographic biofeedback of SMR and beta for treatment of attention deficit disorders in a clinical setting. *Biofeedback and Self-Regulation*, *9*(1), 1–23. https://doi.org/10.1007/BF00998842

Mandryk, R. L., Dielschneider, S., Kalyn, M. R., Bertram, C. P., Gaetz, M., Doucette, A., … Keiver, K. (2013). Games As Neurofeedback Training for Children with FASD. In *Proceedings of the 12th International Conference on Interaction Design and Children* (pp. 165–172). New York, NY, USA: ACM. https://doi.org/10.1145/2485760.2485762

Masasomeha. (2017, January 17). ThinkGear Serial Stream Guide. Retrieved December 4, 2016, from http://developer.neurosky.com/docs/doku.php?id=thinkgear_communications_protocol

McDonald, S. L. (1974). *Feedback Based Eeg Alpha Training and Associated Changes in Task Focus in Grade School Children.* (Ph.D.). The Fielding Institute, United States -- California. Retrieved from http://search.proquest.com.proxy.lib.sfu.ca/pqdtglobal/docview/302741950/citation/9CD66994D8A74DD7PQ/1

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology*, *82*(1), 60.

Mikicin, M., & Kowalczyk, M. (2015). Audio-Visual and Autogenic Relaxation Alter Amplitude of Alpha EEG Band, Causing Improvements in Mental Work Performance in Athletes. *Applied Psychophysiology and Biofeedback*, *40*(3), 219–227. https://doi.org/10.1007/s10484-015-9290-0

Moorcroft, W. H. (2013). What is Sleep and How it is Scientifically Measured. In *Understanding Sleep and Dreaming* (pp. 17–37). Springer US. https://doi.org/10.1007/978-1-4614-6467-9_2

Moore, N. C. (2000). A Review of EEG Biofeedback Treatment of Anxiety Disorders. *Clinical EEG and Neuroscience*, *31*(1), 1–6. https://doi.org/10.1177/155005940003100105

Navidi, W. C. (2006). *Statistics for engineers and scientists* (Vol. 2). McGraw-Hill New York. Retrieved from http://www.academia.edu/download/35263586/SOLUCIONARIO_NAVIDI.pdf

Oberfeld, D., & Franke, T. (2013). Evaluating the robustness of repeated measures analyses: The case of small sample sizes and nonnormal data. *Behavior Research Methods*, *45*(3), 792–812. https://doi.org/10.3758/s13428-012-0281-2

Ogrim, G., Kropotov, J., & Hestad, K. (2012). The quantitative EEG theta/beta ratio in attention deficit/hyperactivity disorder and normal controls: Sensitivity, specificity, and behavioral correlates. *Psychiatry Research*, *198*(3), 482–488. https://doi.org/10.1016/j.psychres.2011.12.041

Osborne, J. W., & Amy, Overbay. (2004). The power of outliers (and why researchers should ALWAYS check for them). Presented at the Practical assessment, research & evaluation 9.6 1-12. Retrieved from https://scholar.google.com/scholar?q=The+power+of+outliers+%28and+why+researchers+should+ALWAYS+check+for+them%29&btnG=&hl=en&as_sdt=0%2C48

Prinsloo, G. E., Derman, W. E., Lambert, M. I., & Rauch, H. L. (2013). The effect of a single session of short duration biofeedback-induced deep breathing on measures of heart rate variability during laboratory-induced cognitive stress: a pilot study. *Applied Psychophysiology and Biofeedback*, *38*(2), 81–90.

Putman, J. (2000). The Effects of Brief, Eyes-Open Alpha Brain Wave Training with Audio and Video Relaxation Induction on the EEG of 77 Army Reservists. *Journal of Neurotherapy*, *4*(1), 17–28. https://doi.org/10.1300/J184v04n01_03

Rebolledo-Mendez, G., Dunwell, I., Martínez-Mirón, E. A., Vargas-Cerdán, M. D., De Freitas, S., Liarokapis, F., & García-Gaona, A. R. (2009). Assessing neurosky's usability to detect attention levels in an assessment exercise. In *Human-Computer Interaction. New Trends* (pp. 149–158). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-02574-7_17

Robbins, R., & Stonehill, M. (2014). *Investigating the NeuroSky MindWave™ EEG headset - Technische Informationsbibliothek (TIB)*. Transport Research Foundation. Retrieved from https://www.tib.eu/en/search/download/?tx_tibsearch_search%5Bdocid%5D=TIBKAT%3A820343749&tx_tibsearch_search%5Bsearchspace%5D=tn&cHash=32f4a0f1822967bf7fffc0484b11263a

Rogers, J. M., Johnstone, S. J., Aminov, A., Donnelly, J., & Wilson, P. H. (2016). Test-retest reliability of a single-channel, wireless EEG system. *International Journal of Psychophysiology*, *106*, 87–96. https://doi.org/10.1016/j.ijpsycho.2016.06.006

Roth, W.-M. (2002). From action to discourse: The bridging function of gestures. *Cognitive Systems Research*, *3*(3), 535–554.

Russell-Chapin, L. R.-C., Kemmerly, T., Liu, W.-C., Zagardo, M. T., Chapin, T., Dailey, D., & Dinh, D. (2013). The Effects of Neurofeedback in the Default Mode Network: Pilot Study Results of Medicated Children with ADHD. *Journal of Neurotherapy*, *17*(1), 35–42. https://doi.org/10.1080/10874208.2013.759017

Sałabun, W. (2014). Processing and spectral analysis of the raw EEG signal from the MindWave. *ResearchGate*, *90*(2), 169–174. https://doi.org/10.12915/pe.2014.02.44

Searle, A., & Kirkup, L. (2000). A direct comparison of wet, dry and insulating bioelectric recording electrodes. *Physiological Measurement*, *21*(2), 271. https://doi.org/10.1088/0967-3334/21/2/307

Shin, D., Lee, G., Shin, D., & Shin, D. (2014). Mental State Measurement System Using EEG Analysis. *SpringerLink*, 451–456. https://doi.org/10.1007/978-3-642-55038-6_70

Simkin, D. R., Thatcher, R. W., & Lubar, J. (2014). Quantitative EEG and Neurofeedback in Children and Adolescents. *Child and Adolescent Psychiatric Clinics of North America*, *23*(3), 427–464. https://doi.org/10.1016/j.chc.2014.03.001

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., … Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *The BMJ*, *338*. https://doi.org/10.1136/bmj.b2393

Stinson, B., & Arthur, D. (2013). A novel EEG for alpha brain state training, neurobiofeedback and behavior change. *Complementary Therapies in Clinical Practice*, *19*(3), 114–118.

Takahashi, J., Yasumura, A., Nakagawa, E., & Inagaki, M. (2014). Changes in negative and positive EEG shifts during slow cortical potential training in children with attention-deficit/hyperactivity disorder: a preliminary investigation. *NeuroReport*, *25*(8), 618–624. https://doi.org/10.1097/WNR.0000000000000156

Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement Science Review*, *2*(2), 1–11.

Thomas, K. P., Vinod, A. P., & Guan, C. (2013). Design of an online EEG based neurofeedback game for enhancing attention and memory. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 433–436). https://doi.org/10.1109/EMBC.2013.6609529

Vollebregt, M. A., van Dongen-Boomsma, M., Buitelaar, J. K., & Slaats-Willemse, D. (2014). Does EEG-neurofeedback improve neurocognitive functioning in children with attention-deficit/hyperactivity disorder? A systematic review and a double-blind placebo-controlled study. *Journal of Child Psychology and Psychiatry*, *55*(5), 460–472. https://doi.org/10.1111/jcpp.12143

Vourvopoulos, A., & Liarokapis, F. (2012). Robot Navigation Using Brain-Computer Interfaces. In *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)* (pp. 1785–1792). https://doi.org/10.1109/TrustCom.2012.247

Vourvopoulos, Athanasios, & Liarokapis, F. (2014). Evaluation of commercial brain–computer interfaces in real and virtual world environment: A pilot study. *Computers & Electrical Engineering*, *40*(2), 714–729. https://doi.org/10.1016/j.compeleceng.2013.10.009

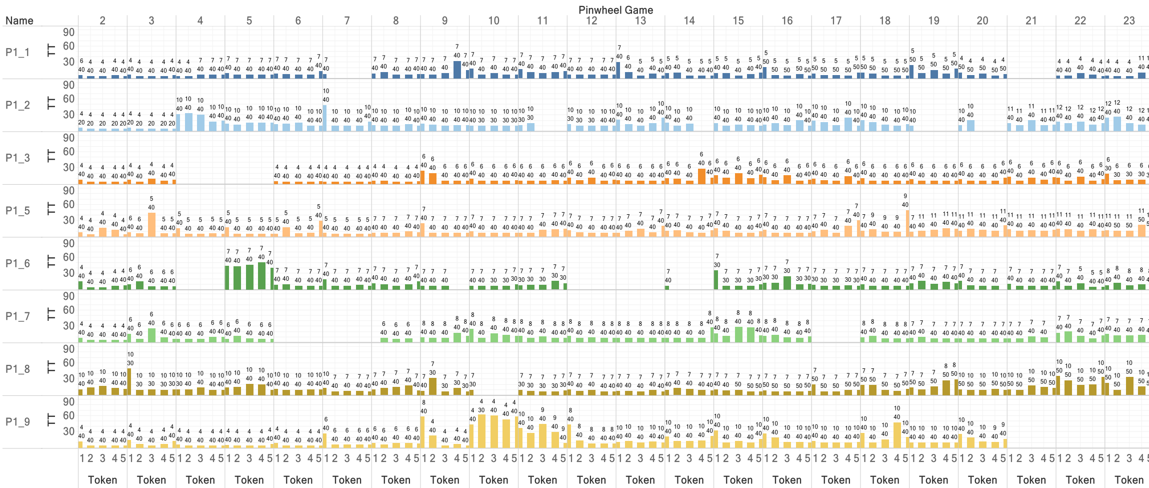
Walker, S. P., Wachs, T. D., Gardner, J. M., Lozoff, B., Wasserman, G. A., Pollitt, E., & Carter, J. A. (2007). Child development: risk factors for adverse outcomes in developing countries. *The Lancet*, *369*(9556), 145–157. https://doi.org/10.1016/S0140-6736(07)60076-2

Zuberer, A., Brandeis, D., & Drechsler, R. (2015). Are Treatment Effects of Neurofeedback Training in Children with ADHD Related to the Successful Regulation of Brain Activity? A Review on the Learning of Regulation of Brain activity and a Contribution to the Discussion on Specificity. Frontiers in Human Neuroscience, 9. https://doi.org/10.3389/fnhum.2015.00135

Appendix A. Data Plots

Time Taken to Collect a Token

Pinwheel Game



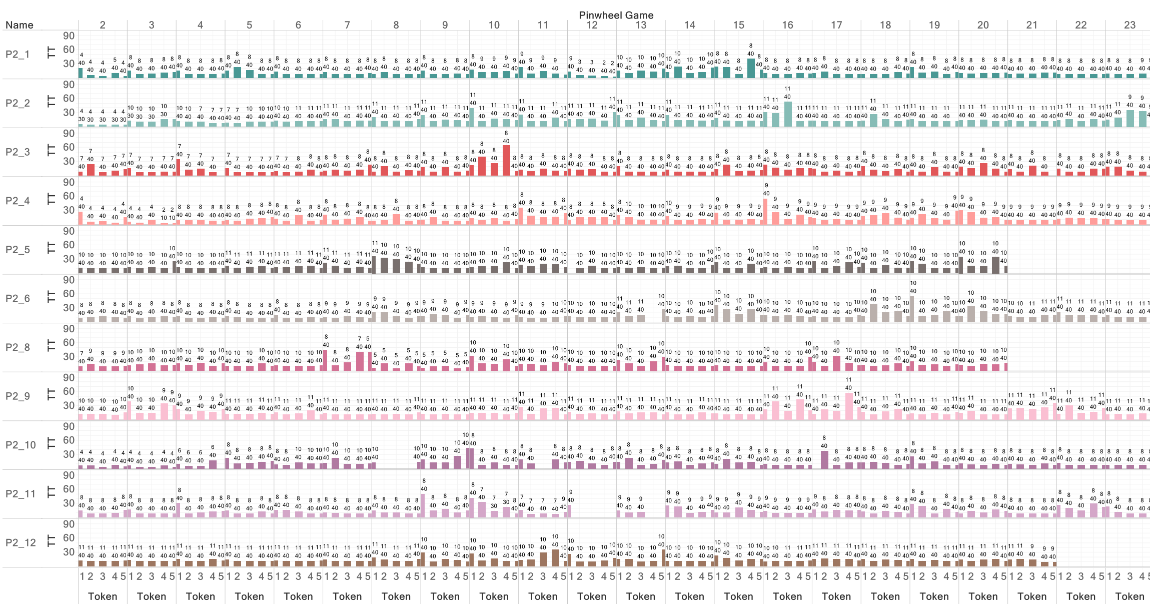


Figure A1 – Individual plot: time taken to collect each token for each participant for 22 sessions (S2 to S23)

Paraglider Game

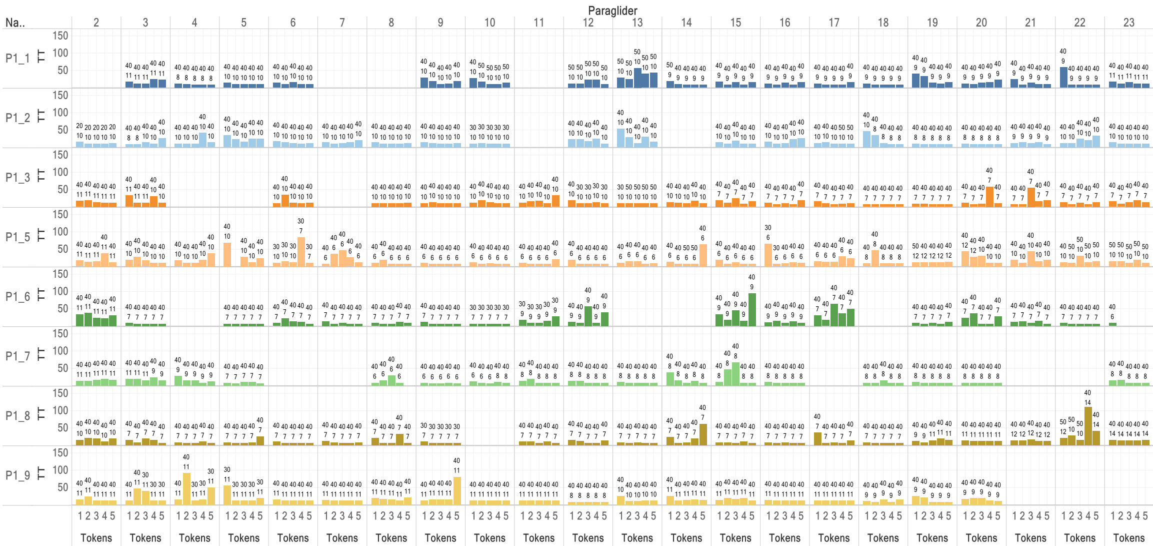
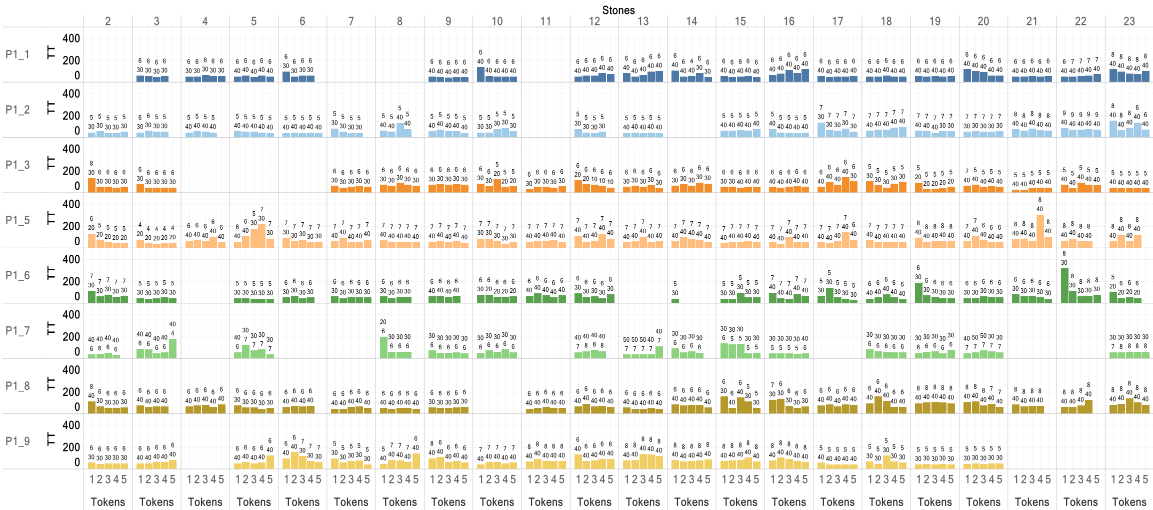


Figure A2 – Individual plot for paraglider game: time taken to collect each token for each participant for 22 sessions (S2 to S23)

Stones Game



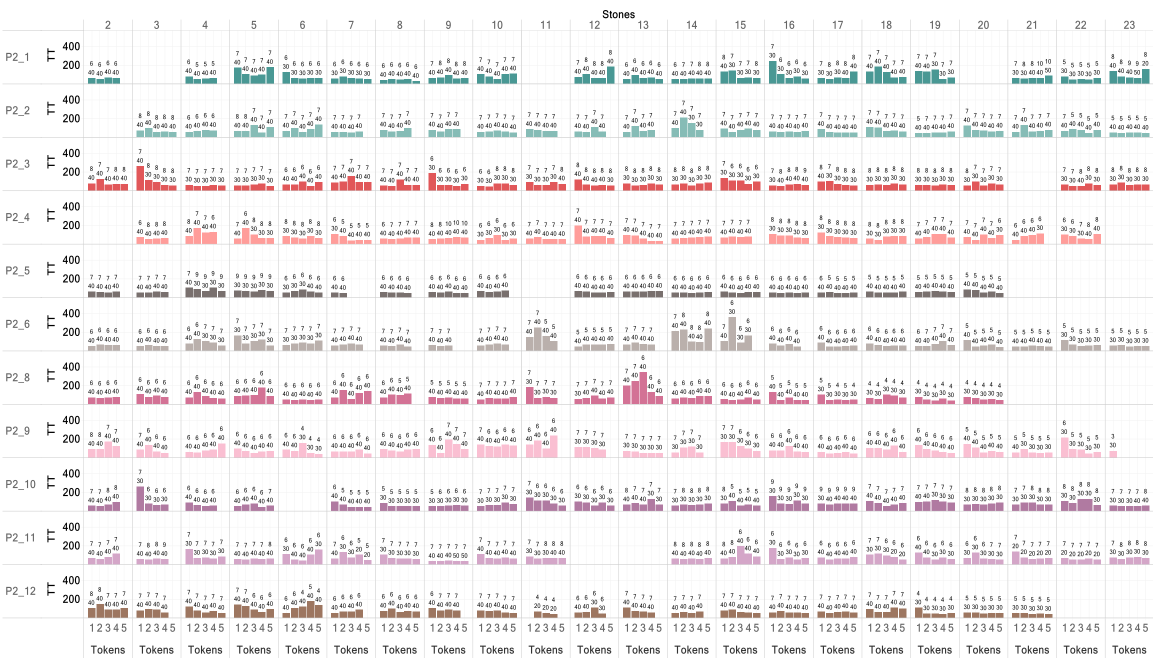


Figure A3 – Individual plot for stones game: time taken to collect each token for each participant for 22 sessions (S2 to S23)

Appendix B.

Ethics Information

A.B.1 Ethics for the Headset Validation Study

Quoted: As is it from Panel of Research Ethics, Government of Canada website[31]

"Article 2.4 REB review is not required for research that relies exclusively on secondary use of anonymous information, or anonymous human biological materials, so long as the process of data linkage or recording or dissemination of results does not generate identifiable information.

Article 2.5 Quality assurance and quality improvement studies, program evaluation activities, and performance reviews, or testing within normal educational requirements when used exclusively for assessment, management or improvement purposes, do not constitute research for the purposes of this Policy, and do not fall within the scope of REB review.

Application Article 2.5 refers to assessments of the performance of an organization or its employees or students, within the mandate of the organization, or according to the terms and conditions of employment or training. Those activities are normally administered in the ordinary course of the operation of an organization where participation is required, for example, as a condition of employment in the case of staff performance reviews, or an evaluation in the course of academic or professional training. Other examples include student course evaluations, or data collection for internal or external organizational reports. Such activities do not normally follow the consent procedures outlined in this Policy.If data are collected for the purposes of such activities but later proposed for research purposes, it would be considered secondary use of information not originally intended for research, and at that time may require REB review in accordance with this

[31] *http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/chapter2-chapitre2/#ch2_en_a2.5*

Policy. Refer to Section D of Chapter 5 for guidance concerning secondary use of identifiable information for research purposes.".

A.B.2 Ethics for the Mind-Full Study

SFU OFFICE OF RESEARCH ETHICS

Street Address	Mailing Address	Director 778.782.6593
Simon Fraser University	8888 University Drive	Associate Director 778.782.9631
Discovery 2	Discovery 2	Manager 778.782.3447
Room 230, 8900 Nelson Way	Burnaby, BC Canada	(removed email)
Burnaby, BC Canada V5A 4W9	V5A 1S6	http://www.sfu.ca/ore.html

Annual Renewal Approval

Study Number: 2014s0540
Study Title: Mind-Full Tablet Games Study - Nepal House Society

Annual Renewal Date: October 23, 2015 **Expiry Date:** October 23, 2016
Principal Investigator: Antle, Alissa **Supervisor:** n/a

SFU Position: Faculty

Faculty/Department: School of Interactive Arts and Technology

SFU Collaborator: n/a

External Collaborator: n/a

Research Personnel: Levisohn, Aaron; Warren, Jillian; Tan, Perry; Cramer, Emily; Eckersley, Rachael; Kirshnamachari Sridharan, Srilekha; Matkin, Brendan; Fan, Min

Funding Source: GRAND NCE

Grant Title: GRAND KIDZ

Documents Approved in this Application:

- Annual Renewal/Progress Report Form
- Annual renewal approval until October 23, 2016

The approval for this study expires on the **Expiry Date**. **Failure to submit an annual renewal form will lead to your study being suspended and potentially terminated.** If you intend to continue your protocol to collect data past the term of approval, you must submit an annual renewal/progress report at least 4 weeks before the expiry date at (removed personal email)

Please notify the Office of Research Ethics at (removed personal email) once you have completed the data collection portion of your project so that we can close the file.

This Notification of Status is your official Annual Renewal Approval documentation for this project. Please keep this document for reference purposes.

Sincerely,

Holly Longstaff, PhD

Acting Associate Director,
Office of Research Ethics.

Appendix C.

Data Pre-Processing Scripts

For each participant:

(1) The data files game_data.csv and headset_data.csv were combined using the script consolidate.py. After combining the files, the game_type (e.g., pinwheel) and the session number were manually added to each row to make it easier for filtering the data.

(2) The GQSP.java was used to calculate the GQSP for each session

(3) The AverageRA.java was used to calculate the mean relaxation per session for pinwheel and paraglider game and the mean attention per session for the stones game.

(4) The AverageEEG.java was used to calculate the mean EEG amplitudes per session for all the three games.

(5) To normalize the time taken to collect tokens, MS Excel was used. The number of tokens in every session was manually checked using the individual plots (Appendix B).

Refer supplementary materials for more information on the data pre-processing scripts.