VALIDATION OF A LOW-COST MOBILE EEG DEVICE FOR STUDYING LANGUAGE-RELATED ERP COMPONENTS

by

Hannah Begue Hayes

A thesis presented in partial fulfillment of the requirements for the
Degree of Master of Arts in Psychology

Middle Tennessee State University
May 2024

Thesis Committee:

Dr. Cyrille Magne, Chair

Dr. James Houston

Dr. William Langston

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Dr. Cyrille Magne, for accepting me into his lab after I became academically homeless, as well as for his patience in the face of many panicked emails.

I would like to thank (the future Dr.) Stephanie Powell-Wolfe for her keen perspective and for keeping me grounded in matters of both life and learning.

Finally, and most importantly, I'd like to thank my husband, Conner Hayes, for many late-night proofreads, a never-ending supply of fresh coffee, and undying support in every way. Thank you for making my dreams part of your own.

ABSTRACT

InteraXon's Muse 2 is a commercial EEG headband designed to provide neurofeedback during mindfulness exercises. The Muse's low-cost, ease of use, and mobility provide an attractive alternative to the limitations associated with a traditional wired EEG system. The goal of this study was to validate the Muse for conducting language ERP (event-related potential) research using a Semantic Relatedness Judgment Task (SRJT) and a Stroop Task. The Muse measured distinct N1, P2, and N400 components in both tasks. Monte Carlo permutation tests showed significant differences between Match and Mismatch conditions in the SRJT but not the Stroop Task. These results indicate that the Muse can measure language-related ERPs in certain contexts and supports the implementation of the Muse headband for future ERP studies of language. Potential benefits of the Muse include increasing sample size and diversity and democratizing ERP research. Further research is needed to refine the parameters that determine the ideal range of use for the Muse EEG device in ERP studies.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

Introduction

The event-related potential (ERP) technique measures neural responses to physical stimuli using electroencephalography (EEG) (Luck, 2014). Stimuli presentation and participant responses are time-locked to EEG data to determine the relationship between changes in cortical voltage, physical stimuli, and behavior. ERPs refer to local minimums and maximums called components, e.g. the P300 or N400, that occur a predictable amount of time after presentation of a specific type of stimulus. Typical ERP studies compare the components of two or more conditions to determine how neural responses differ based on variations in stimuli. Due to the temporal accuracy and continuous recording capabilities of EEGs, ERP studies are highly effective for covertly observing multiple phases of cognitive processes (Luck, 2014).

Typical EEG devices require a wired connection to a computer, which limits ERP research. Traditional wired research EEG systems are cumbersome, pricey, and stationary; data collection occurs almost exclusively in a well-equipped laboratory and can only be performed by trained researchers (Casson, 2019). These limitations result from the prioritization of signal quality in device design, e.g., to maximize signal-to-noise ratio, minimize transient artefacts, and ensure temporally accurate data transmission. Recent advances in EEG development have produced a diverse array of wireless systems designed for clinical, research, and consumer applications. Each system has a unique combination of portability, electrode type and configuration, size, and wireless protocols (Niso et al., 2023). Mobile EEG prices range from less than $1,000 to over $50,000 (Niso et al., 2023). InteraXon’s Muse 2 stands out for its simplicity, portability, and relatively low cost.

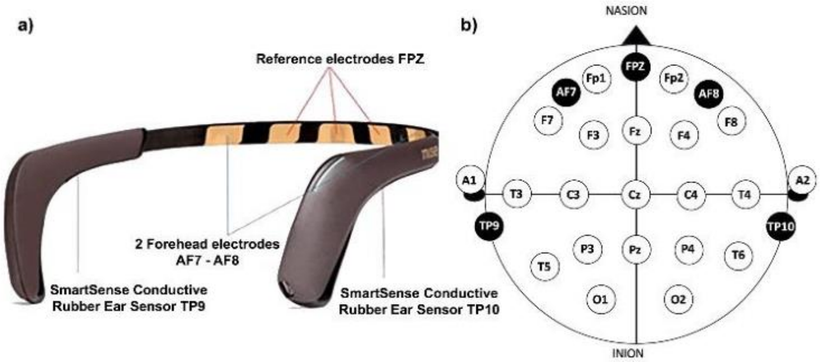
InteraXon's Muse 2

InteraXon's Muse 2 is a dry electrode, four channel EEG headband designed and manufactured for commercial use. The Canadian manufacturer markets the Muse as a neurofeedback tool for guided meditation. The Muse connects to InteraXon's mobile application via Bluetooth and provides neurofeedback by incorporating EEG-based auditory cues into guided meditations (InteraXon Inc., 2021). While the Muse's effectiveness as a stress reduction tool remains undetermined (Svetlov et al., 2019), many features that attract consumers also appeal to ERP researchers. The mobile EEG device is unobtrusive, easy to use, and costs approximately $250 (InteraXon Inc., 2021). As an essential research benefit, several open-source software functionalities have been developed in the scientific community to stream raw EEG signal to outside applications for further processing and analysis.

The Muse 2 headband rests across the middle of the forehead and fits over the ears like eyeglasses (See Figure 1a). The locations of the Muse's fixed electrodes are analogous to the electrode locations of the 10-20 International system used to map electrode placement in the nets of wired EEG systems (Milnik, 2006). The device is fitted with two sensors behind the ears (TP9 and TP10), one electrode on each side of the forehead (AF7 and AF8), and a reference electrode in the center of the forehead (FPz) (See Figure 1b) (Mansi et al., 2021; Niso et al., 2023).

Figure 1

Muse 2 Electrode Placement



From Mansi, S. A., Pigliautile, I., Porcaro, C., Pisello, A. L., & Arnesano, M. (2021). Application of wearable EEG sensors for indoor thermal comfort measurements. *Acta IMEKO (2012), 10*(4), 214. https://10.21014/acta_imeko.v10i4.1180

The Muse's dry electrodes use pressure to form connection rather than traditional conductive gel or electrolytic water, which can be unpleasant and off-putting for study participants. Dry electrode systems are simpler and easier to use but risk decreased signal, increased noise, and decreased power when compared to gel and wet electrode systems (Mathewson et al., 2016; Niso et al., 2022). Despite their limitations, Kam et al. (2019) suggest dry electrodes are nearing wet systems in effectiveness.

The Muse connects locally to devices via Bluetooth and is compatible with the Lab-Streaming Layer (LSL), an open-source ecosystem designed to facilitate streaming and synchronization of data recording (Niso et al., 2023). The LSL synchronizes raw EEG data from the Muse with behavioral and stimulus presentation data.

The Muse has previously been implemented in a variety of research areas including mindfulness and stress reduction, mental and emotional state classification, dementia pain management, substance use, and lie detection (Gillani et al., 2021; Hawley et al., 2021; Bird et al., 2018; Nanthini et al., 2021; Pu et al., 2021; Saengmolee et al., 2022; Youssef et al., 2018). This literature supports validation of the Muse as a research method and neurofeedback intervention tool in the listed capacities, but ERP literature using the Muse is limited to date. Krigolson et al. (2017, 2021) reported distinct N200, P300, and reward positivity components measured during oddball and reward learning tasks using the Muse. A team of media researchers measured distinct ERP components during an affective image processing task but found no ERP differences between the two conditions (Jahn et al., 2022). Fickling et al. (2020) suggest that the electrode configuration of the Muse may be insufficient for ERP measurement. Due to the abundance of literature on non-ERP applications of the Muse and the sparse literature on its ERP applications, this study aims to determine if the Muse has a place as an accurate and reliable ERP methodology.

Validation of the Muse as a tool to measure language ERPs offers the opportunity to address questions in language research historically limited by traditional EEG devices. Researchers can expand data collection beyond the laboratory to improve generalizability and increase statistical power through larger, more diverse samples. This flexibility will increase access to specific populations such as children, those with learning disabilities, and those who otherwise might be unwilling or unable to travel to the laboratory.

Language-Related ERPs

To be implemented as a research tool in language ERP studies, the Muse must be able to measure the ERP components most studied in language, particularly the N400. The N400 effect is a negativity occurring 300 to 500 ms post-stimulus that is greater for unexpected than for expected stimuli. Although initially discovered in language research, the N400 effect has been measured with mathematic and pictorial stimuli, indicating that the component is produced by the processing of meaning rather than the processing of language (Rabovsky, 2023). In language research, an incongruent or unexpected word produces a greater negativity than a congruent or expected word. For example, the sentence “I put my dirty clothes in the laundry *potato*” creates a much greater negativity after the final word than if the sentence reads “I put my dirty clothes in the laundry *hamper”* (Kutas and Hillyard, 1980). Because the N400 is the product of a wide range of lexical and semantic modulators, thorough understanding of this component is crucial to broader understanding of language processes (Rabovsky, 2023).

To date, no ERP studies on language have been published using the Muse 2. Wireless EEG systems have been successfully implemented in language ERP research to measure the N400 effect, but the gel 32-and 24-electrode systems used in these studies present a different set of assets and limitations than the Muse (Park and Donaldson, 2019; Fjaellingsdal et al., 2016). To determine the Muse’s language ERP capabilities, this study replicated two tasks which demonstrate well-known paradigms for language-related ERPs, a Stroop Task and a Semantic Relatedness Judgment Task (SRJT). In both tasks, an N400 component is expected in response to incongruent stimuli (Heidlmayr et al., 2020; Kutas and Federmeier, 2011).

Semantic Relatedness Judgment Task

In the Semantic Relatedness Judgment Task, the participant is presented with a sequence of prime and target word pairs that are either related (e.g. apple-banana) or unrelated (e.g. cup-banana). The participant responds that the pair is related or unrelated (Koivisto and Revonsuo, 2001). The target with an incongruent prime produces a heightened negativity 300-600 ms after presentation of the second word when compared to targets with a congruent prime (Kutas and Federmeier, 2011). The elevated amplitude for the N400 after incongruent pairs is a result of the incorrect expectation formed by the prime (Rabovsky, 2023) and is associated with an increase in time between stimulus presentation and participant response (Koivisto and Revonsuo, 2001).

Stroop Task

In a Stroop Task, participants are presented with a color word (blue, green, red) whose text is blue, green, or red. In the congruent condition, the color of the text matches the color the word represents. In the incongruent condition, the color of the text is different than the color the word represents (Stroop, 1935). The participant must respond with the color of the text rather than the word itself (Sahinoglu and Dogan, 2016). The contradiction between color naming and the more automatic process of word reading produces interference between intentional and automatic processes, resulting in multiple distinct components (Sahinoglu and Dogan, 2016; Heidlmayr et al., 2020).

Incongruent stimuli (e.g. the word “red” in green text) are expected to produce a heightened early negativity 200-500 ms after stimulus presentation and a late sustained positivity 500-800 ms post-presentation compared to congruent stimuli (Heidlmayr et al., 2020). The negativity in the Stroop is analogous to the semantically produced N400 in timing and both engage common features of cognitive control. However, the Stroop N400 originates from interference rather than semantic processing (Applebaum et al., 2009; Heidlmayr et al., 2020). In the Stroop Task, the late positivity is associated with conflict processing (Coderre et al., 2011; Donohue et al., 2016) and semantic processing (Applebaum et al., 2009). The incongruent condition produces a greater positivity than the congruent condition (Heidlmayr et al., 2020).

Present Study

This study thus aims to confirm that the Muse can accurately measure language-related ERPs, specifically the N400, in a Stroop Task and a Semantic Relatedness Judgment Task.

Based on the existent literature and its deficits, the hypotheses for the current study were as follows:

1. For Muse EEG data recorded during the Stroop Task, incongruent stimuli would produce an early negativity 200-500 ms post stimulus onset followed by a late positivity 500 to 800 ms post stimulus onset.
2. For Muse EEG data recorded during the Semantic Relatedness Judgment Task, word pairs with an incongruent prime would produce a negativity at 300 to 600 ms following target word onset of greater amplitude than that of a congruent target and prime.

CHAPTER II

Method

Participants

Thirty-seven participants (18 males, 19 females) participated in the study. Thirty-four of the participants were undergraduate students recruited through the MTSU psychology research pool. Three additional participants were recruited by word of mouth. Thirty participants identified as White, three as Black/African-American, one as Asian, and three as Other. Three participants identified as Hispanic and/or Latino. The average participant age was 19.97 years (SD = 3.77). Participants recruited through the research pool received course credit, but otherwise no compensation was provided. This study was approved by the Institutional Review Board, and written consent was obtained from each participant prior to initiating the study. Exclusionary criteria eliminated minors under 18 years of age and participants with blindness, inhibiting visual deficits, and hearing impairment. One participant met exclusionary criteria due to reported deficiency in color vision, so their data was not included in analysis of the Stroop Task.

Procedure

Ideally, ERP studies are conducted in a soundproof room, but a less restrictive setting was chosen to mimic the settings of future studies more closely (Luck et al., 2014). This study was performed in a quiet, but not sound-proof, laboratory with the researcher seated near the participant. The participant signed the informed consent form upon arrival and sat down in front of the monitor used for the upcoming tasks. The participant then donned the Muse based on researcher instruction. Participants were given the option to place the Muse on themselves or to allow the researcher to do it. After initiating the first task, the researcher adjusted the Muse based on the live EEG visualizer. After completing both tasks, the participant removed the Muse, and the researcher debriefed them.

Language Tasks

The stimuli for the Stroop Task and the SRJT were developed using the PsychoPy Builder interface, a Python-based graphical user interface for stimulus design. Each participant completed both tasks. Task order was counterbalanced across participants to mitigate priming and fatigue effects.

Stroop Task

The Stroop Task presented a color word (red, blue, green) with colored text (red, blue, or green). Stimuli was presented at a visual angle of approximately four degrees. The interstimulus interval was variable based on response times. The conditions were Match (text matches word color) and Mismatch (text color is different than letter color.) The instructions informed participants to ignore the word and press a keyboard arrow corresponding to text color: left for red, down for blue, and right for green. Participants performed six practice trials comprised of matching word/color stimuli before continuing to the actual task. The participants responded with the text color to proceed to the next trial. The task contained 90 trials composed of 45 match and 45 mismatch stimuli randomly presented within one block. Participants spent approximately five minutes completing the task. Participants were not provided verbal instructions during the task to avoid participants correcting from selecting the word to the color mid-task.

Semantic Relatedness Judgment Task

The SRJT was formatted similarly to the Stroop Task, and stimuli was similarly presented at a visual angle of approximately four degrees. The interstimulus interval between prime and target words was fixed at one second. The interstimulus interval between word pairs was variable based on answer times. The instructions informed participants to press “R” for related and “U” for unrelated based on their judgment of the presented word pair. Participants then completed eight practice trials containing word pairs from both the Match and Mismatch conditions. The prime word was presented for 500 ms followed by 500 ms of blank screen before target word presentation. The target word remained on the screen until the participant responded. The task was comprised of 112 random trials containing 56 pairs of related words and 56 pairs of unrelated words. After the first 56 trials, participants took a short break in their seats and proceeded at their own pace. The task lasted approximately seven minutes. Two sets of word pairs were counterbalanced among participants. Word pairs were previously validated to produce an N400 effect (Holcomb and Neville, 1990). The researcher provided verbal instructions upon request.

EEG Data Collection

ERP research requires the ability to time-lock precise EEG recording with stimulus presentation. PsychoPy was selected to create the stimuli based on its ability to interface with the Muse via the Lab-Streaming Layer (Peirce et al., 2019; *LabStreamingLayer*, 2023). PsychoPy features the option to insert individual lines of Python code in the Builder GUI. Custom code was added to each task to calculate the exact screen refresh rate and to monitor each frame. Additional code was inserted into the tasks to send triggers to the LSL to insert timestamped markers representing stimulus presentation and condition into the EEG data. EEG data recording began at the start of each task. See https://osf.io/42pzj/?view_only=09d03df54b0546e587ea29b0cfa60d9a to view the code used for the two tasks.

The study was performed with a Windows laptop facing the participant and an external monitor and keyboard facing the researcher. The open-source BlueMuse software connected the Muse to the laptop via Bluetooth (Richard et al., 2019). BlueMuse streamed the raw EEG data to Lab Streamer. The Lab Streaming Layer synchronized the incoming data streams from BlueMuse and PsychoPy into a single time-locked ".xdf" file. Participants performed the tasks on the laptop while the researcher viewed their screen to monitor EEG channels and correct recording.

EEG Data Processing

Data processing was performed in MATLAB (MathWorks, Inc.), the EEGLAB toolbox (Delorme and Makeig, 2004), and the FieldTrip Open-Source Toolbox (Oostenveld et al., 2011). First, the raw EEG signal was high-pass filtered at 0.5 Hz to minimize slow drift and low-pass filtered at 20 Hz to reduce excess noise. While raw EEG signal from a wired device is typically high-pass filtered at 0.1 Hz (Luck, 2014), 0.5 Hz was selected as a high-pass filter based on observation of low frequency noise in raw Muse data. N400 studies generally use sensors placed on the mastoids (i.e., part of the skull right behind the ears) as references (Li et al., 2018). The data was thus re-referenced offline to the average of the two temporo-parietal sensos (TP9 and TP10) due to their close proximity to the mastoids. Artefact Subspace Reconstruction removed transient artefacts by referencing a clean segment of data and removing segments of with greater than 40 standard deviations of variance. Epochs were created by extracting the time frame 100 ms before the stimulus marker to 900 ms after. The marker occurred with the target word in the SRJT and with the colored word in the Stroop Task. The window for baseline correction was -100 ms to 0 ms. Only correct trials were used in final analysis.

CHAPTER III

Results

Behavioral Data

Table 1 contains descriptive statistics for accuracy rates and response times during the Semantic Relatedness Judgment Task and the Stroop Task. Comparisons of average response times and accuracy rates for the within-subject fixed factor of condition (Match, Mismatch) for the Stroop and Semantic Related Judgment Tasks were conducted using the MIXED procedure in SAS Studio (version 3.80, REML estimation, Kenward and Roger Method 2 degrees of freedom) with a familywise alpha of .05. Response times were only calculated for correct responses. For the Stroop Task, response times were shorter for the match condition than the mismatch condition, $F(1, 2963) = 73.09$, $p < .001$, and accuracy rates were higher for the match than the mismatch condition, $F(1,3238) = 239.56$, $p < .001$. For the SRJT, response times were shorter for the match condition than for the mismatch condition, $F(1,2017) = 120.57$, $p < .001$. Accuracy rates were not significantly different between conditions in the SRJT.

Table 1
Descriptive Statistics for All Variables

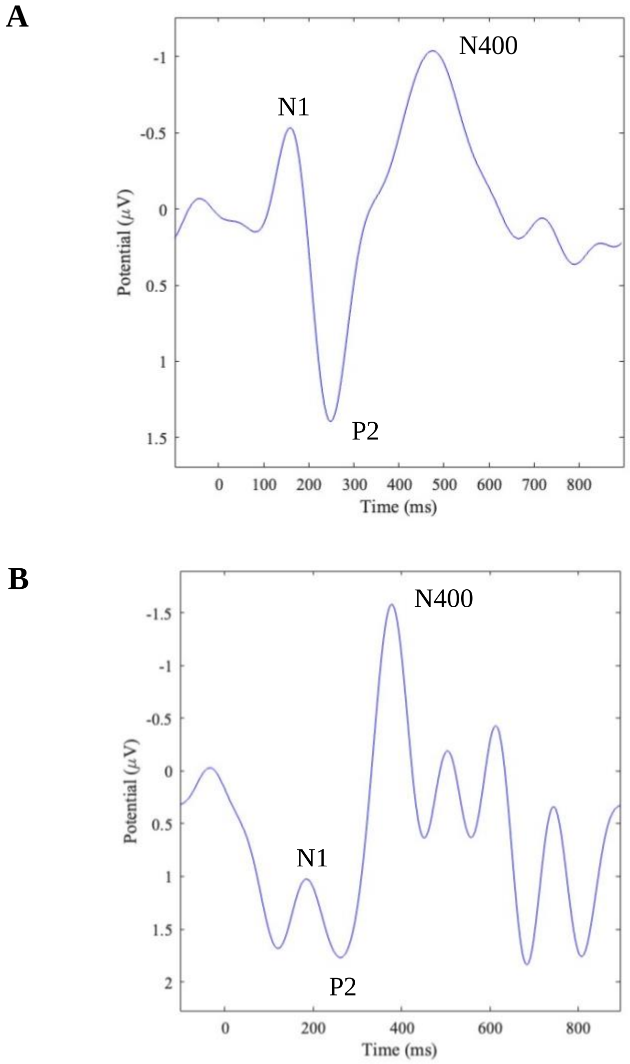
		SRJT (*N*=37)		Stroop Task (*N*=36)	
Variable		M	SD	M	SD
Match	Accuracy Rate (%)	97.06	0.17	98.83	0.12
	Response Time (ms)	828.04	0.38	730.56	0.30
Mismatch	Accuracy Rate (%)	97.64	0.98	84.20	0.36
	Response Time (s)	1047.22	0.51	842.16	0.39
Total	Accuracy Rate (%)	97.35	0.16	91.51	0.28
	Response Time (s)	937.79	0.46	781.90	0.35

EEG Data

The grand average ERPs for the combined match and mismatch conditions are depicted in Figure 2 for both the SRJT and Stroop Task. A visual inspection of the waveforms revealed distinct N1, P2, and N400 components in both tasks.

Figure 2

Grand average ERP waveforms for all conditions for the (A) Semantic Relatedness Judgement Task and (B) Stroop Task.

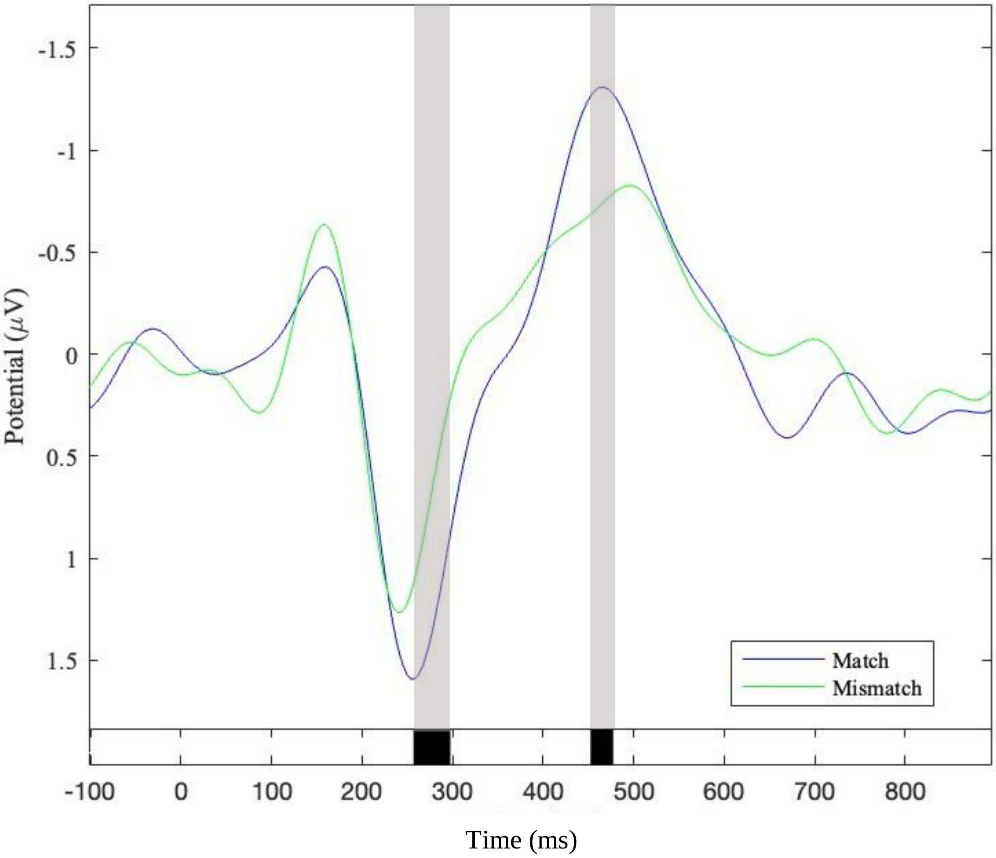


Semantic Relatedness Judgment Task

Four participants' data were removed due to elevated noise levels based on visual inspection. No participant had accuracy levels lower than the threshold of 80%, so there were no rejections due to low accuracy. Monte Carlo permutation tests indicated two periods of significant difference between the Match and Mismatch conditions at an alpha of .05 (see Figure 3). Between 257.8 ms and 296.9 ms ($p < .05$), the Mismatch condition was significantly more negative than the Match condition. Between 453.1 ms and 476.6 ms ($p < .05$), the Mismatch condition showed an increased positivity compared to the Match condition.

Figure 3

Grand average ERP waveforms of match versus mismatch for the Semantic Relatedness Judgment Task.

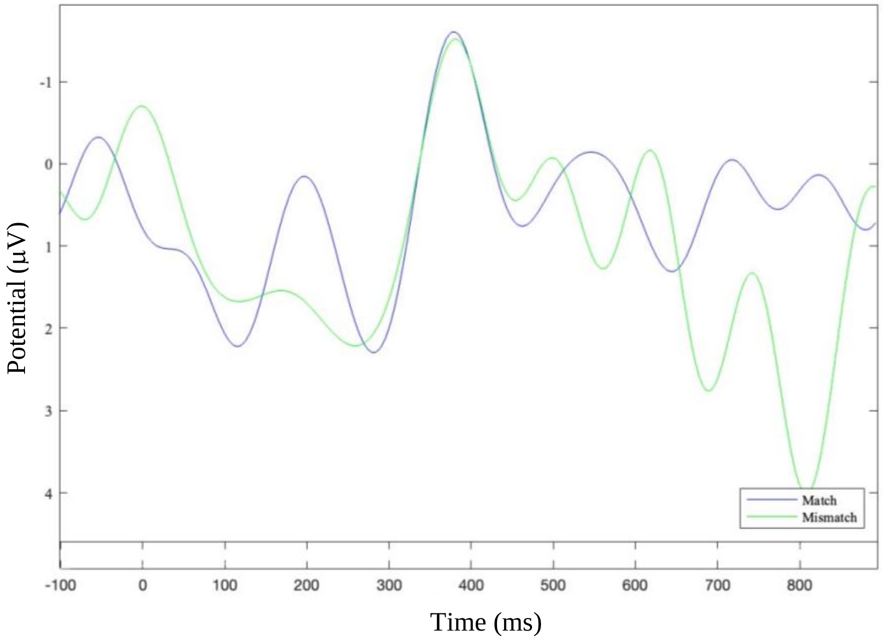


Stroop Task

Five participants' data were eliminated due to accuracy rates less than 60%. Two additional participants were removed due to elevated noise levels based on visual inspection. Monte Carlo permutation tests indicated no periods of significant difference between the Match and Mismatch conditions at an alpha of .05 (See Figure 4).

Figure 4

Grand average ERP waveforms of match versus mismatch for the Stroop Task.



CHAPTER IV

Discussion

For the Semantic Relatedness Judgment Task, the longer average response time of the Mismatch condition compared to the Match condition is in congruence with previous findings, e.g., Koivisto and Revonsuo (2001). For the Stroop Task, the lower average accuracy rate and longer average response time of the Mismatch condition compared to the Match condition also reflects previous findings (Stroop, 1935; Heidlmayr et al., 2020).

Visual inspection of the Muse ERP data measured in both the Stroop and SRJT indicate the presence of N1, P2, and N400 components. The N1 is an early negativity occurring at approximately 100 ms post-stimulus, and the P2 is a positivity at approximately 200 ms post-stimulus (Luck, 2024). The N1 and P2 components occur in tandem to form the N1/P2 complex, which is associated with lower-level orthographic processing including extraction of visual features and word-form access (Pokhoday et al., 2023). The N1 also reflects spatial attention (Luck, 2014). The N400 is associated with semantic processing and lexical access (Kutas & Federmeier, 2011). While the N1/P2 complex is not as frequent a variable of interest in language research as the N400, the presence of N1 and P2 components indicate the Muse can measure a range of different components.

The significant differences in ERP components measured during the Semantic Relatedness Judgment Task support the hypothesis that the Mismatch condition would produce a larger N400 between 300 ms and 600 ms than the Match condition, thus strongly supporting the validity of using the Muse EEG headband as a tool for language ERP research. Contrary to the hypothesis, the ERP data recorded during the Stroop Task did not show evidence that incongruent data would lead to an early negativity followed by a late positivity. Specifically, the

Match and Mismatch conditions did not exhibit significant differences at any point. Several limitations associated with the study’s design, including technical constraints related to the Muse device, may account for this discrepancy.

Limitations

Despite ample evidence of ERP differences during Stroop Tasks measured with wired EEG systems (Heidlmayr et al., 2020), data collected using the Muse showed no significant differences between the Match and Mismatch conditions. Failure to support the hypothesis that the Stroop Task would produce an early negativity and late positivity is the result of both technological limitations of the device itself and failure to predict extraneous variables. Studies using a wired EEG system for ERP studies during Stroop Tasks generally utilize more trials than this study, e.g. Rebai et al. (1997) and Applebaum et al. (2019). This study lacked neutral stimuli (e.g., grey words, words that do not represent colors, or colored non-word symbols) as controls (Rebai et al., 1997; Heidlmayr et al., 2020). Additionally, the researcher only provided instructions prior to the task, and five participants’ accuracy rates were below 60%, indicating they may have reported the word rather than the color. Furthermore, the data recorded during the Stroop Task was much noisier than the data for the SRJT, which could be attributed to participants moving their eyes and muscles to see the color labels on the keys.

In summary, the Muse did not reveal the expected ERP differences between the Match and Mismatch condition in the Stroop Task, highlighting the importance of considering experimental design and technical factors. Future studies using the Muse should increase the number of trials for each condition, provide sufficient instruction to participants, and minimize artefacts. Artefacts are created by eye blinks, eye movement, and muscle movement (Luck et al., 2014). Examples of factors that would reduce artefacts are adding a fixation cross between stimuli (Applebaum et al., 2009) and positioning response keys to minimize eye movement. Replication of the SRJT with these changes in mind would likely produce stronger evidence of N400 differences than was measured in this study.

Future Implications

Muse data collection was straightforward and efficient. One researcher independently collected data for 37 participants in one week. The study took no more than 30 minutes to conduct with less than five minutes of set-up. Every piece of equipment used to collect data fit inside one laptop bag. The Muse 2 costs $250, uses only free software, and can be easily ordered online. The device's portability, efficiency, and affordability offer flexibility in data collection not possible with a traditional wired EEG system.

A potential benefit of the Muse is the opportunity to reach previously inaccessible populations. Studies conducted in a university laboratory are often limited to undergraduate students seeking class credit or members of the community seeking compensation. Restricting studies to these demographics limits the scope of research on literacy, learning disabilities, and development, creating a significant barrier to generalizability. A researcher using the Muse can collect data at a school and reach a desirable sample size of a difficult to reach population in a relatively short period of time. In addition to schools, the Muse studies can be conducted at locations with greater access to target and marginalized populations such as literacy centers and English as a Second Language programs.

In addition to expanding traditional study demographics, the Muse also democratizes language ERP research. Researchers without the funding necessary to purchase a traditional EEG system can use the Muse as an effective alternative. This can offer early-career researchers in fields with less funding for neuroscience research, such as education, the ability to conduct language ERP research. Similarly, the Muse offers a straightforward alternative for students who lack the experience or the timeframe to independently conduct EEG studies with a wired system.

The Muse poses solutions to common EEG challenges of equity and inclusion. Because the sensors contact hairless locations, the Muse can easily be donned by individuals with protective hairstyles or dreadlocks. Participants do not have to worry about wetting their hair or removing conductive gel. The Muse is less invasive, more comfortable, and easier to adjust compared to a traditional EEG net or cap, which makes the headband preferable for individuals with sensory issues or those averse to touch.

Incorporating the Muse into interventions and diagnostics can aid in language learning, education, and literacy. In individuals learning a second language, neural correlates of language learning can be measured before behavioral changes occur (McLaughlin et al., 1999). In children diagnosed with developmental dyslexia, ERPs recorded during a reading task were significantly different pre-and post-intervention between the group whose performance improved and the group whose did not (Hasko et al, 2014). In both cases, neural differences could be identified in ERPs prior to the identification of behavioral differences. By using the Muse to easily record ERPs, educators can not only identify learning differences early but can use that information to guide intervention and improve outcomes on the individual level.

Conclusion

InteraXon’s Muse 2 offers an inexpensive, mobile alternative to traditional wired EEG systems for use in language research. The Muse demonstrated the ability to measure significant N400 differences during the Semantic Relatedness Judgment Task. Although the device did not measure the predicted ERP differences during the Stroop Task, changes in experimental design have been identified to improve outcomes for future studies. Validation of the Muse will significantly impact seasoned and novice researchers alike and possesses the potential to revolutionize language education.

REFERENCES

Appelbaum, L. G., Meyerhoff, K. L., & Woldorff, M. G. (2009). Priming and Backward Influences in the Human Brain: Processing Interactions during the Stroop Interference Effect. *Cerebral Cortex*, *19*(11), 2508–2521. https://doi.org/10.1093/cercor/bhp036

Bird, J. J., Manso, L. J., Ribeiro, E. P., Ekart, A., & Faria, D. R. (2018). A study on mental state classification using EEG-based brain-machine interface. *2018 International Conference on Intelligent Systems (IS), Intelligent Systems (IS), 2018 International Conference On*, , 795-800. https://10.1109/IS.2018.8710576

Casson, A. (2019). Wearable EEG and beyond. *Biomedical Engineering Letters*, *9*(1), 53-71. https://10.1007/s13534-018-00093-6

Coderre, E., Conklin, K., & Van Heuven, W. J. B. (2011). Electrophysiological measures of conflict detection and resolution in the Stroop task. *Brain Research*, *1413*, 51–59. https://doi.org/10.1016/j.brainres.2011.07.017

Donohue, S. E., Appelbaum, L. G., McKay, C. C., & Woldorff, M. G. (2016). The neural dynamics of stimulus and response conflict processing as a function of response complexity and task demands. *Neuropsychologia*, *84*, 14–28. https://doi.org/10.1016/j.neuropsychologia.2016.01.035

Fickling, S. D., Bollinger, F. H., Gurm, S., Pawlowski, G., Liu, C. C., Hajra, S. G., Song, X., & D'Arcy, R. C. N. (2020). Distant sensor prediction of event-related potentials. *IEEE Transactions on Biomedical Engineering*, *67*(10), 2916-2924. https://10.1109/TBME.2020.2973617

Fjaellingsdal, T. G., Bleichner, M. G., Ruigendijk, E., & Scherbaum, S. (2016). The N400 effect during speaker-switch-towards a conversational approach of measuring neural correlates of language. *Frontiers in Psychology, 7*https://10.3389/fpsyg.2016.01854

Gillani, S. F., Umar Saeed, S. M., Ali, Z. U. A. E. D., Monis, Shabbir, Z., & Habib, F. (2021). Prediction of perceived stress scores using low-channel electroencephalography headband. *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST), Applied Sciences and Technologies (IBCAST), 2021 International Bhurban Conference On,* , 401-406. https://10.1109/IBCAST51254.2021.9393175

Hawley, L. L., Rector, N. A., DaSilva, A., Laposa, J. M., & Richter, M. A. (2021). Technology supported mindfulness for obsessive compulsive disorder: Self-reported mindfulness and EEG correlates of mind wandering. *Behaviour Research and Therapy, 136*https://10.1016/j.brat.2020.103757

Hasko, S., Groth, K., Bruder, J., Bartling, J., & Schulte-Körne, G. (2014). What does the brain of children with developmental dyslexia tell us about reading improvement? ERP evidence from an intervention study. *Frontiers in Human Neuroscience*, *8*. APA PsycInfo. https://doi.org/10.3389/fnhum.2014.00441

Heidlmayr, K., Kihlstedt, M., & Isel, F. (2020). A review on the electroencephalography markers of Stroop executive control processes. *Brain and Cognition, 146*, 105637. https://10.1016/j.bandc.2020.105637

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and cognitive processes*, *5*(4), 281-312.

InteraXon Inc. (2021, *Muse: The brain sensing headband.* choosemuse.com

Jahn, N. T., Meshi, D., Bente, G., & Schmälzle, R. (2022). Media neuroscience on a shoestring: Examining electrocortical responses to visual stimuli via mobile EEG. *Journal of Media Psychology: Theories, Methods, and Applications,* https://10.1027/1864-1105/a000348

Kam, J. W. Y., Griffin, S., Shen, A., Patel, S., Hinrichs, H., Heinze, H., Deouell, L. Y., & Knight, R. T. (2019). Systematic comparison between a wireless EEG system with dry electrodes and a wired EEG system with wet electrodes. *NeuroImage, 184*, 119-129. https://10.1016/j.neuroimage.2018.09.012

Krigolson, O. E., Williams, C. C., Norton, A., Hassall, C. D., & Colino, F. L. (2017). Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. *Frontiers in Neuroscience, 11*https://10.3389/fnins.2017.00109

Krigolson, O. E., Hammerstrom, M. R., Abimbola, W., Trska, R., Wright, B. W., Hecker, K. G., & Binsted, G. (2021). Using muse: Rapid mobile assessment of brain performance. *Frontiers in Neuroscience, 15*https://10.3389/fnins.2021.634147

Koivisto, M., & Revonsuo, A. (2001). Cognitive representations underlying the N400 priming effect. *Brain Research. Cognitive Brain Research, 12*(3), 487-490. https://10.1016/S0926-6410(01)00069-6

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*https://10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect

Semantic Incongruity. *Science*, *207*(4427), 203–205. http://www.jstor.org/stable/1683915

LabStreamingLayer. (2023) [Computer software]. Swartz Center for Computational Neuroscience. https://github.com/sccn/labstreaminglayer

Li, Y., Wang, Y., Zhang, B., Wang, Y., & Zhou, X. (2018). Electrophysiological Responses to Expectancy Violations in Semantic and Gambling Tasks: A Comparison of Different EEG Reference Approaches. *Frontiers in Neuroscience*, *12*, 169. https://doi.org/10.3389/fnins.2018.00169

Luck, S. J. (2014). *An introduction to the event-related potential technique, second edition* (2nd ed.). MIT Press.

Mansi, S. A., Pigliautile, I., Porcaro, C., Pisello, A. L., & Arnesano, M. (2021). Application of wearable EEG sensors for indoor thermal comfort measurements. *Acta IMEKO (2012)*, *10*(4), 214. https://10.21014/acta_imeko.v10i4.1180

Mathewson, K. E., Harrison, T. J. L., & Kizuk, S. A. D. (2017). High and dry? comparing active dry EEG electrodes to active and passive wet electrodes. *Psychophysiology*, *54*(1), 74-82. https://10.1111/psyp.12536

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703–704. https://doi.org/10.1038/nn1264

Nanthini, K., Pyingkodi, M., Sivabalaselvamani, D., & Kaviya. (2022). EEG signal analysis for emotional classification. *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Electronics and Sustainable Communication Systems (ICESC), 2022 3rd International Conference On*, , 192-198. https://10.1109/ICESC54411.2022.9885391

Niso, G., Romero, E., Moreau, J. T., Araujo, A., & Krol, L. R. (2023). Wireless EEG: A survey of systems and studies. *NeuroImage (Orlando, Fla.)*, *269*, 119774. https://10.1016/j.neuroimage.2022.119774

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Park, J. L., & Donaldson, D. I. (2019). Detecting the neural correlates of episodic memory with mobile EEG: Recollecting objects in the real world. *NeuroImage*, *193*, 1-9. https://10.1016/j.neuroimage.2019.03.013

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pokhoday, M., Bermúdez-Margaretto, B., Malyshevskaya, A., Kotrelev, P., Shtyrov, Y., & Myachykov, A. (2023). Eye-Tracking Methods in Psycholinguistics. In M. Grimaldi, E. Brattico, & Y. Shtyrov (Eds.), *Language Electrified* (Vol. 202, pp. 731–752). Springer US. https://doi.org/10.1007/978-1-0716-3263-5_23

Pu, L., Lion, K. M., Todorovic, M., & Moyle, W. (2021). Portable EEG monitoring for older adults with dementia and chronic pain—A feasibility study. *Geriatric Nursing*, *42*(1), 124–128. Supplemental Index.

Rebai, M., Bernard, C., & Lannou, J. (1997). The Stroop's test evokes a negative brain potential, the N400. *International Journal of Neuroscience*, *91*(1–2), 85–94. APA PsycInfo. https://doi.org/10.3109/00207459708986367

Rabovsky, M. (2023). Neural Underpinnings of Semantic Processing. In M. Grimaldi, E. Brattico, & Y. Shtyrov (Eds.), *Language Electrified* (Vol. 202, pp. 527–550). Springer US. https://doi.org/10.1007/978-1-0716-3263-5_16

Richard, J., Bacon, R., Blaizot, J., Boissier, S., Boselli, A., NicolasBouché, Brinchmann, J., Castro, N., Ciesla, L., Crowther, P., Daddi, E., Dreizler, S., Duc, P.-A., Elbaz, D., Epinat, B., Evans, C., Fossati, M., Fumagalli, M., Garcia, M., … Yang, B. (2019). *BlueMUSE: Project Overview and Science Cases*. https://doi.org/10.48550/ARXIV.1906.01657

Saengmolee, W., Chaisaen, R., Autthasan, P., Rungsilp, C., Sa-Ih, N., Cheaha, D., Kumarnsit, E., & Wilaiprasitporn, T. (2022). Consumer-grade brain measuring sensor in people with long-term kratom consumption. *IEEE Sensors Journal, 22*(6), 6088-6097. https://10.1109/JSEN.2022.3147207

Sahinoglu, B., & Dogan, G. (2016). Event-related potentials and the stroop effect/olay-iliskili potansiyeller ve stroop etkisi. *The Eurasian Journal of Medicine, 48*(1), 53. https://10.5152/eurasianjmed.2016.16012

Stroop, J. R. (1992). Studies of interference in serial verbal reactions (reprinted from journal experimental-psychology, vol 18, pg 643-662, 1935). *Journal of Experimental Psychology-General, 121*(1), 15-23. https://search.ebscohost.com/login.aspx?direct=true&db=edswss&AN=A1992HE76400004&site=eds-live&scope=site

Svetlov, A. S., Nelson, M. M., Antonenko, P. D., Joseph P.H. McNamara, & Bussing, R. (2019). Commercial mindfulness aid does not aid short-term stress reduction compared to unassisted relaxation. *Heliyon, 5*(3)https://10.1016/j.heliyon.2019.e01351

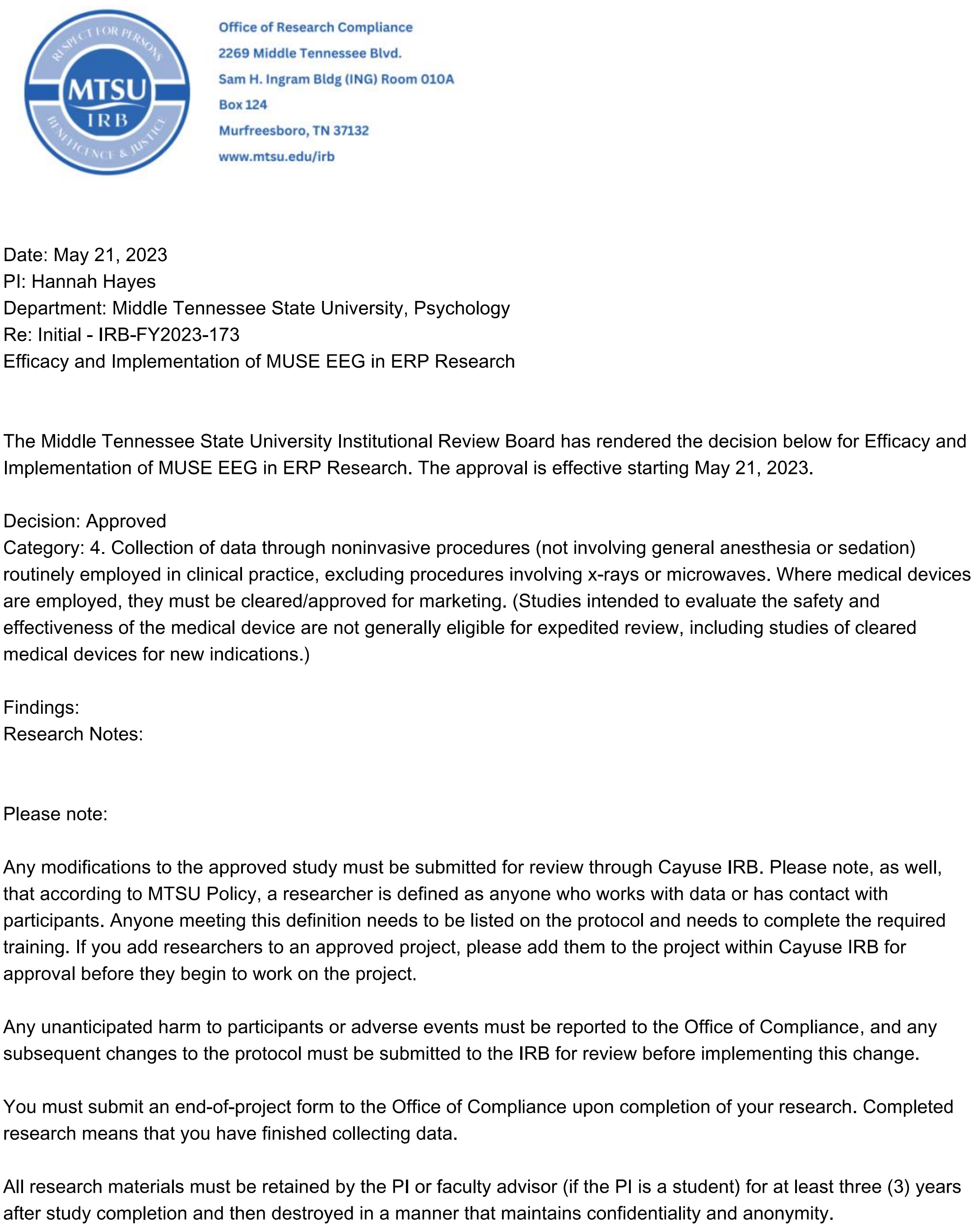
Wascher, E., Heppner, H., & Hoffmann, S. (2014). Towards the measurement of event-related EEG activity in real-life working environments. *International Journal of Psychophysiology*, *91*(1), 3-9. https://10.1016/j.ijpsycho.2013.10.006

Youssef, A. E., Ouda, H. T., & Azab, M. (2018). MUSE: A portable cost-efficient lie detector. *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2018 IEEE 9th Annual,* , 242-246. https://10.1109/IEMCON.2018.8614795

APPENDIX

APPENDIX A

IRB Approval Letter



Office of Research Compliance
2269 Middle Tennessee Blvd.
Sam H. Ingram Bldg (ING) Room 010A
Box 124
Murfreesboro, TN 37132
www.mtsu.edu/irb

Date: May 21, 2023
PI: Hannah Hayes
Department: Middle Tennessee State University, Psychology
Re: Initial - IRB-FY2023-173
Efficacy and Implementation of MUSE EEG in ERP Research

The Middle Tennessee State University Institutional Review Board has rendered the decision below for Efficacy and Implementation of MUSE EEG in ERP Research. The approval is effective starting May 21, 2023.

Decision: Approved
Category: 4. Collection of data through noninvasive procedures (not involving general anesthesia or sedation) routinely employed in clinical practice, excluding procedures involving x-rays or microwaves. Where medical devices are employed, they must be cleared/approved for marketing. (Studies intended to evaluate the safety and effectiveness of the medical device are not generally eligible for expedited review, including studies of cleared medical devices for new indications.)

Findings:
Research Notes:

Please note:

Any modifications to the approved study must be submitted for review through Cayuse IRB. Please note, as well, that according to MTSU Policy, a researcher is defined as anyone who works with data or has contact with participants. Anyone meeting this definition needs to be listed on the protocol and needs to complete the required training. If you add researchers to an approved project, please add them to the project within Cayuse IRB for approval before they begin to work on the project.

Any unanticipated harm to participants or adverse events must be reported to the Office of Compliance, and any subsequent changes to the protocol must be submitted to the IRB for review before implementing this change.

You must submit an end-of-project form to the Office of Compliance upon completion of your research. Completed research means that you have finished collecting data.

All research materials must be retained by the PI or faculty advisor (if the PI is a student) for at least three (3) years after study completion and then destroyed in a manner that maintains confidentiality and anonymity.

All approval letters and study documents are located within the Study Details in Cayuse IRB.

We wish you a successful research project,

Middle Tennessee State University Institutional Review Board

APPENDIX B

Stimuli Used in Semantic Relatedness Judgment Task

RELATED	UNRELATED	TARGET
VENT	HOUSE	AIR
GOOD	SOIL	BAD
KEG	COMB	BEER
COMB	LOOSE	BRUSH
CHILL	WIN	COLD
MOM	COME	DAD
SOIL	KEG	DIRT
SKETCH	STAND	DRAW
YOLK	FALSE	EGG
BRAWL	MIST	FIGHT
BLAZE	YOLK	FIRE
MIST	BROTH	FOG
PAL	RIP	FRIEND
COME	DAY	GO
LOW	PAL	HIGH
HOUSE	CORK	HOME
WIN	TALL	LOSE
DAY	INK	NIGHT
INK	MOM	PEN
TUG	SKETCH	PULL
KING	VENT	QUEEN
FRIGHT	EAST	SCARE
YELL	TUG	SCREAM
TALL	FRIGHT	SHORT
STAND	THICK	SIT
BROTH	YELL	SOUP
HALT	LOW	STOP
RIP	KING	TEAR
THICK	HALT	THIN
LOOSE	CHILL	TIGHT
FALSE	BRAWL	TRUE
EAST	GOOD	WEST
CORK	BLAZE	WINE
SUBTRACT	PEDAL	ADD
SALOON	ACRE	BAR
PEDAL	ICING	BIKE

FRACTURE	HANDBAG	BREAK
CONSTRUCT	SELECT	BUILD
ICING	AIRPORT	CAKE
ALTER	SALOON	CHANGE
SELECT	BUTCHER	CHOOSE
OPEN	ODOR	CLOSE
SOFA	SUBTRACT	COUCH
REPAIR	OBSERVE	FIX
EMPTY	GIGGLE	FULL
ASSIST	EMPTY	HELP
ACRE	SOFA	LAND
GIGGLE	CONSTRUCT	LAUGH
NOISY	REPAIR	LOUD
BUTCHER	OPEN	MEAT
AIRPORT	INSTRUCT	PLANE
HANDBAG	NOISY	PURSE
BOULDER	ASSIST	ROCK
ODOR	ALTER	SMELL
INSTRUCT	FRACTURE	TEACH
OBSERVE	BOULDER	WATCH
FRONT	SAND	BACK
SAND	SHRUB	BEACH
GIRL	DUSK	BOY
SHRUB	LAST	BUSH
CHEF	ITCH	COOK
WEEP	GRASP	CRY
DUSK	BRIDE	DAWN
KNOB	JOG	DOOR
DINE	SCALE	EAT
TOUCH	SHOVE	FEEL
SEEK	SPOON	FIND
LAST	DINE	FIRST
SPOON	GIRL	FORK
FUEL	CLOCK	GAS
BRIDE	WEEP	GROOM
GRASP	RICH	HOLD
LEAP	NEW	JUMP
BLEND	FRONT	MIX
NEW	SEEK	OLD
RICH	BREEZE	POOR
SHOVE	QUIZ	PUSH

JOG	TOUCH	RUN
ITCH	TOSS	SCRATCH
LOOK	FUEL	SEE
ILL	BEST	SICK
NAP	LOOK	SLEEP
NORTH	CHEF	SOUTH
QUIZ	KNOB	TEST
TOSS	ILL	THROW
CLOCK	NORTH	TIME
SCALE	BLEND	WEIGHT
BREEZE	NAP	WIND
BEST	LEAP	WORST
REQUEST	TINY	ASK
MATTRESS	CASHEW	BED
CARDBOARD	SPOTLESS	BOX
BEETLE	ATTEMPT	BUG
PURCHASE	SPICY	BUY
TABLE	LABOR	CHAIR
CHEDDAR	VACATE	CHEESE
SPOTLESS	REQUEST	CLEAN
STUMBLE	PURCHASE	FALL
BANNER	LISTEN	FLAG
LISTEN	FLAVOR	HEAR
SPICY	STUMBLE	HOT
TARDY	MATTRESS	LATE
VACATE	DISCUSS	LEAVE
ANGRY	CHEDDAR	MAD
CASHEW	CHLORINE	NUT
CHLORINE	ANGRY	POOL
UNCOOKED	TARDY	RAW
TINY	UNCOOKED	SMALL
DISCUSS	BEETLE	TALK
FLAVOR	BANNER	TASTE
ATTEMPT	CARDBOARD	TRY
LABOR	TABLE	WORK

APPENDIX C

Stimuli Used in Stroop Task

BLUE	BLUE	BLUE
RED	RED	RED
GREEN	GREEN	GREEN