**ABSTRACT**

MCGOUGH, OLIVIA H. Examining the Effects of Signal Words and Motivation on Vigilance. (Under the direction of Dr. Christopher B. Mayhorn).

Many high-stakes jobs require strong attentional control and high vigilance capabilities. As such, understanding factors that influence vigilance is essential to the design of user-centered protocols. Because empirical research has shown that sustained attention tasks are associated with vigilance decrements over time, two experiments were conducted to explore various mitigation strategies that may improve performance outcomes.

Experiment One assessed the differences in signal word variations on sustained attention task accuracy, reaction time (RT), and boredom. One hundred seventy-one participants were randomly assigned to complete a vigilance task that varied between-subjects in seven signal word/text highlight color groups (e.g., WARNING, CAUTION). Participants in the experimental conditions also received a motivational message outlining the consequences of noncompliance of the task in addition to the signal word variation, that read "[SIGNAL WORD]: If you do not complete the following task with at least 95% accuracy, you will be required to re-start the activity." Participants completed self-report questionnaires to assess trait and state boredom levels, and all participants completed a modified version of the Mackworth Clock Test, an experimental paradigm designed to test vigilance over time. Results showed significant accuracy differences between the control group and several of the signal word/text highlight color groups, but no accuracy differences were found between the signal word groups themselves. Additionally, significantly faster RTs were seen in participants who received the motivational message in conjunction with a signal word and alert symbol compared to participants who received no message prior to completing the vigilance task. No significant state boredom differences were found between groups after completing the vigilance task when controlling for trait boredom. However, results showed significant associations between trait/state boredom and vigilance task accuracy, such that higher boredom was related to reduced accuracy on the task. Further, faster RTs were associated with a greater number of correct detections on the task.

Experiment Two complemented Experiment One by exploring the impact of several socially motivating stimuli on sustained attention performance. One hundred sixty-three participants were randomly assigned to four groups that used different socially motivating competition formats (e.g., control, gamification leaderboard, reward incentive, combined). Participants completed the same questionnaires and task as Experiment One, designed to assess trait/state boredom levels and performance on a basic vigilance task. Results showed partial support that reward incentivization may be a more effective competition format to promote accuracy on sustained attention tasks compared to ranked high score leaderboards. No significant state differences were found between the different competition formats after completing the vigilance task when controlling for trait boredom. However, similar to the results of Experiment One, significant associations were found between trait/state boredom and task accuracy. Additionally, Experiment Two indicated that faster RTs were associated with a greater number of correct detections on the vigilance task, akin to the findings in Experiment One.

Overall, both studies provided insight into the utility of various strategies to improve performance on vigilance-based tasks. Additionally, both studies highlighted the importance of considering boredom as an individual factor that can impact performance on sustained attention tasks. This information may be helpful in the consideration of design choices for various user protocols surrounding tasks that require strong vigilance capabilities.

*Keywords*: Vigilance, sustained attention, motivation, competition, boredom

Examining the Effects of Signal Words and Motivation on Vigilance

by
Olivia H. McGough

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Psychology

Raleigh, North Carolina
2025

APPROVED BY:

| | |
|---|---|
| Dr. Christopher B. Mayhorn<br>Committee Chair | Dr. Jing Feng |
| Dr. Anna Behler | Dr. Yingchen He |

## DEDICATION

To my mother – your unconditional support and confidence in me helped me achieve this goal and continues to motivate me to pursue my biggest dreams. I love and miss you always.

To my father – thank you for being my greatest role model. You’ve taught me kindness, strength, and resilience, and I am so proud to be your daughter.

To Jake – my incredible partner and very best friend, thank you for being my rock. I’m so grateful for the life that we’ve built together, and I’m so excited for what’s to come.

## BIOGRAPHY

Olivia McGough is a Psychology Ph.D. candidate at North Carolina State University. Olivia graduated from Rowan University in Spring 2021, where she received a Bachelor of Science degree in Psychological Science. She began her doctoral program at NC State in Fall 2021, pursuing her Ph.D. with a concentration in Human Factors and Applied Cognition. During her graduate school career, Olivia had the opportunity to work as a Human Factors Specialist with User-View Inc. She also participated in the Federal Aviation Administration's (FAA) Gateways Student Internship program for five summers as a Human Factors Intern, beginning in Summer 2020. In July 2024, Olivia accepted a full-time position with the FAA's Human Systems Integration Branch at the William J. Hughes Technical Center for Advanced Aerospace, where she works as an Engineering Research Psychologist.

**ACKNOWLEDGEMENTS**

Thank you to my committee members: Dr. Anna Behler, Dr. Yingchen He, and Dr. Jing Feng for your support throughout the dissertation process and my greater graduate school journey. I truly appreciate your time, efforts, and insightful feedback. I'd like to especially thank my advisor, Dr. Chris Mayhorn, for sharing your wisdom and expertise with me. It was an honor to learn from you over the past four years.

Thank you to Addison Davis for your contributions to the literature review in support of this research. Your assistance was extremely helpful, and I am grateful for all your hard work.

To Dr. Janey Barnes and User-View Inc., thank you for the opportunity to learn and develop professionally under your exceptional guidance and for your support during my graduate school career.

Thank you also to my colleagues at the FAA's Human-Systems Integration Branch for all of your encouragement, support, and guidance during my doctoral program and dissertation preparation.

I would also like to thank Dr. Tom Dinzeo and Dr. Sherry Pujji for first introducing me to the world of psychological research and for supporting my academic and professional development since my undergraduate days.

Finally, thank you to my friends and peers from the HFAC program at NCSU. I'd especially like to acknowledge Dr. Nina Ferreri for being an amazing peer mentor during my graduate school career. I've enjoyed learning with you all over the past few years and will cherish the many fond memories.

## TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

**Examining the Effects of Signal Words and Motivation on Vigilance**

Sustained attention describes the process of maintaining attention and effort over continuous periods (Ko et al., 2017). A related concept is vigilance, or one's ability to sustain attention without distraction (Shaw et al., 2010). Large amounts of research provide evidence for a phenomenon known as the vigilance decrement, where performance declines over time (Davies & Parasuraman, 1982; Jerison, 1970; Mackworth, 1948; See et al., 1995; Warm, 1977; Warm et al., 2008).

Two primary theories are commonly used to explain the vigilance decrement: Mindlessness Theory (Manly et al., 1999; Robertson et al., 1997) and Resource Theory (Hancock & Warm, 1989; Hockey, 1997; Parasuraman & Moulouna, 1987). Mindlessness Theory states that decrements in performance on sustained attention tasks are due to the monotony of the task that can lead to mind-wandering and boredom, thus increasing errors. Resource Theory, on the other hand, opines that sustained attention tasks are demanding of cognitive resources. Resource Theory also explains that cognitive resources deplete over time, leading to errors (Davies & Parasuraman, 1982; Grier et al., 2003; Hancock & Warm, 1989; Kahneman, 1973; Robison & Nguyen, 2023; Smit et al., 2004; Warm et al., 2008; Wickens, 2002).

In many instances, errors due to a lack of vigilance could result in dangerous outcomes, as many of the most safety-critical jobs rely on human operators. Transportation Security Administration (TSA) airport baggage screening, radiology, and lifeguarding are a few examples of careers where errors due to a decline in vigilance would be detrimental to human safety. TSA agents use screening technologies to aid their efficiency but must visually scan each baggage item to detect contraband. In a similar fashion, radiologists are trained to interpret medical imaging to make accurate diagnoses and treatment recommendations. Likewise, lifeguards must continuously scan an area of water to ensure all swimmers are safe. In each of these roles, there is a large potential for harmful consequences if a mistake occurs. As such, understanding factors that could influence performance on vigilance-based tasks is important to the design of user-centered protocols and technologies. Some factors that may influence sustained attentional performance include (a) signal words/warning systems, (b) boredom, and (c) motivation. In the following sections, each of these factors are addressed, and hypotheses are made about their impact on performance.

**Sustained Attention and Signal Words**

Warning systems are often used in workplace environments to promote employee safety. Warning systems have three intended purposes: to reduce workplace accidents, to explain critical safety-related information to users and anyone in the vicinity, and to encourage safe behaviors (Conzola & Wogalter, 2001). There are three signal words typically used to denote varying levels of hazards, DANGER, WARNING, and CAUTION (ANSI, 2002). Each of these words are used in different contexts to indicate varying hazard levels (ANSI, 2002; FMC Corporation, 1985; Westinghouse Electric Corporation, 1981; Peckham, 2006). According to the ANSI (2002) Z535 standards, DANGER is recommended "when serious injury or death *will* occur if the directive is not followed" (ANSI, 2002; Mayhorn et al., 2015, pp. 345). WARNING is advised "when serious injury or death *may* occur if the directive is not followed" (ANSI, 2002; Mayhorn et al., 2015, pp. 345). Finally, CAUTION is recommended "when less severe personal injuries or property damage may occur if the directive is not followed" (ANSI, 2002; Mayhorn et al., 2015, pp. 345). The ANSI (2002) Z535 standards also delineate text highlight color for each signal word. DANGER should be highlighted in red, WARNING in orange, and CAUTION in yellow

(ANSI, 2002). Further, an alert symbol should be included on each hazard panel, represented by an exclamation point within a triangle. (ANSI, 2002). Signal word messages, or warning systems, should only be used either as a supplement to other safety measures, or as a final resort when it is impossible to design a hazard out of a system (Mayhorn et al., 2015). Designing a hazard away is always the best option when possible. However, including warning systems as an additional layer of protection can be beneficial to draw a user's attention to a safety hazard and as an overall tool to promote safety.

**Sustained Attention, Signal Words, and Motivation**

Broadly speaking, warning systems should include an explanation of the hazard. They should also list instructions for how to avoid the hazardous outcome, as well as any consequences that may occur if someone does not comply with the warning system (Rogers et al., 2000). Awareness, understanding, and response to a warning system can depend on several factors. The Communication-Human Information Processing (C-HIP) model is a framework that can be referenced when understanding the factors that may impact the effectiveness of warning systems (Wogalter, 2006b). The model is divided into two main sections and each section is broken down further into specific stages. The first section's stages depict the communication of the warning message being passed from a source to a receiver. The second section of the model looks at the receiver's cognitive processing of the information presented in the warning message. The model explains that for internal processing to occur in the second section, effective communication of the warning must have occurred in the first section. Once the warning has been received, the second section of the model outlines several stages of processing: attention switch, attention maintenance, comprehension memory, attitude beliefs, motivation, and resulting behavior. The receiver must successfully process information at each stage, or else the information will not continue to the next stage. For example, an individual may not notice a warning message in the first place. Alternatively, they may read the warning but not understand its meaning. Rather, the receiver may have understood the warning's meaning but not believed the message to be true and therefore chosen to ignore it. On the other hand, if the individual believed the message, there may still be an issue with motivation. In this case, the receiver may not have the motivation to perform the behavior that the warning instructed (Mayhorn et al., 2015). According to the C-HIP model, successful processing at each stage will result in compliance with the warning. However, unsuccessful processing at any of the stages will result in noncompliance (Wogalter, 2006b).

Motivation drives an individual to complete a task and is a critical step in the C-HIP model. As described by the Theory of Planned Behavior (Azjen, 1991), motivation bridges an individual's internal attitudes with their behaviors. There are many influences that may impact an individual's motivation to comply with a warning message. For example, an individual may be less motivated to adhere to an instruction if the time and/or effort associated with the instruction is perceived as too much (Mayhorn et al., 2015; Wogalter et al., 1987; Wogalter et al., 1989). On the other hand, awareness of the potential outcomes that could occur due to noncompliance can in themselves be a motivating factor for adherence. For example, high risk for injury if one does not comply with a warning may influence an individual's willingness to follow the instructions, regardless of the perceived undesirable time/effort costs to perform the instructions. Research conducted by Wogalter et al. (1991, 1993, 1999e) found that individuals reported higher levels of willingness to comply with a warning instruction when there was a belief that noncompliance would result in a high likelihood of severe injury.

McGough & Mayhorn (2022) and McGough & Mayhorn (2023) measured accuracy on variations of a sustained attention task in combination with signal words and messaging intended to motivate accurate performance. In McGough & Mayhorn (2022), participants received a written prompt that read, "**WARNING:** If you do not complete the following task with at least **<u>95%</u>** accuracy, you will be required to re-start the activity." Results showed that participants who received this instruction performed significantly more accurately on the task than in the control condition that did not include this instruction.

Based on the ANSI (2002) Z535 signal word recommendations (ANSI, 2002), the written prompt was revised for the subsequent study to read, "**CAUTION:** If you do not complete the following task with at least **<u>95%</u>** accuracy, you will be required to re-start the activity." According to the ANSI (2002) Z535 standards, WARNING panels are advisable when "serious injury or death *may* occur if the directive is not followed," whereas CAUTION panels are advised when "less severe personal injuries or property damage may occur if the directive is not followed" (Mayhorn et al., 2015, pp. 345). As the ANSI (2002) Z535 standards are recommendations for designing safety warnings in the workplace, the signal panel use case recommendations do not easily translate to tasks such as performance incentivization on a laboratory sustained attention paradigm. Therefore, in this particular study, WARNING was interpreted to indicate a high risk of inconvenience due to delay, whereas CAUTION indicated a lower risk and consequence. In this case, where a basic sustained attention task possessed no safety hazard but only a potential risk of needing to restart the task, CAUTION was deemed more appropriate than WARNING based on these definitions.

In McGough & Mayhorn (2022), significant accuracy differences were seen between the experimental group that received a WARNING prompt and the control group, yet no significant differences were seen between groups in a component of the subsequent study (McGough & Mayhorn, 2023) where the same motivational prompt was displayed but the signal word was changed to CAUTION.[1] This difference in findings was unanticipated and prompted additional questions. It is important to note that reliable comparisons cannot be made between the two studies as the design and methodology were different for each. As such, there are numerous potential reasons for the difference in findings, and further research is needed to explore the comparisons in detail. Specifically, the discrepancy in findings warrants research that can compare the impact of signal word choice (WARNING/CAUTION) on the strength of the participants' task performance in response to the motivational nature of the prompt/instructions.

There are differing opinions on whether WARNING and CAUTION and their respective text colors (orange and yellow) cause users to reliably distinguish between the hazards and thus consider their behavior appropriately. Past research has shown that most people are not able to reliably distinguish between hazards that are depicted using orange and yellow (Chapanis, 1994; Mayhorn et al., 2004; Wogalter et al., 1998a). Additionally, previous research provided evidence that users do not reliably differentiate between CAUTION and WARNING in workplace safety hazards (Mayhorn et al., 2015). However, the ANSI (2002) Z535 standards recommend using DANGER, WARNING, or CAUTION to reliably distinguish between varying levels of hazards (ANSI, 2002) in workplace environments. Moreover, it is unclear how signal words and text highlight color impact attention and accuracy specifically on vigilance-based tasks.

In addition to considering the impact of signal word variation on sustained attention performance, it is also important to consider the impact of the message content itself on an

---

[1] Of note, an additional component of McGough & Mayhorn (2023) explored the impact of accuracy feedback on performance. Significantly higher accuracy was seen when participants received trial-by-trial accuracy feedback following each individual trial compared to participants who received no individual trial feedback.

individual's motivation. As previously stated, motivation is a critical step in the C-HIP model, and perception of a warning system can highly influence one's response (Wogalter, 2006b). At the most fundamental level, two types of motivation exist: *intrinsic* motivation and *extrinsic* motivation. As defined by Self-Determination Theory (SDT, Deci & Ryan, 1985), intrinsic motivation is "doing something because it is inherently interesting or enjoyable" and extrinsic motivation is "doing something because it leads to a separable outcome" (Ryan & Deci, 2000, pp. 55). Intrinsic motivation involves engaging in some task or activity because of an internal desire that is not due to any external influence. For example, a person may be intrinsically motivated to spend time with their friends, solely because they enjoy the activity. Extrinsic motivation, on the other hand, is motivation that is influenced by external factors such as the desire to earn a reward or avoid a punishment. Examples of extrinsic motivation include a student preparing well for an exam to receive a good grade, an employee working extra hours to earn overtime pay, or a student doing their homework to not receive detention. In one study, participants were incentivized in a sustained attention task when they were instructed that the task would be shortened if their reaction time stayed within a certain time frame. Results showed that participants in this group displayed less of a vigilance decrement, less decrements in motivation, and less of a decrement in alertness when compared to a group that was not incentivized (Garner et al., 2024). Other research found evidence that providing an incentive as motivation for performance significantly reduced the extent of the vigilance decrement (Robison & Nguyen, 2023). Similarly, the prompt used in McGough & Mayhorn (2022, 2023) was intended to motivate performance accuracy on a sustained attention task. As previously described, the message read, "If you do not complete the following task with at least **<u>95%</u>** accuracy, you will be required to re-start the activity." and included a signal word (WARNING or CAUTION) at the start of the message. McGough & Mayhorn (2022) found evidence that participants who received the incentivizing message performed more accurately on the vigilance task compared to participants who received no message. In McGough & Mayhorn (2023), where the signal word was changed, those findings were not replicated. However, this difference in findings could potentially be attributed to several factors, as the studies had different designs. For example, McGough & Mayhorn (2022) prefaced the motivating message with the signal word "WARNING", which is generally considered to be more severe than the signal word used in McGough & Mayhorn (2023), "CAUTION." While the signal word variation may have altered the perceived severity of the message, it is also necessary to recognize that the message was likely in itself motivating, even without the presence of a signal word. Therefore, one focus of the current study was to determine whether the impact of varying signal words produced performance accuracy differences when presented with the same motivating message. However, extensive literature exists supporting the general idea that extrinsic motivation reduces vigilance decrements (Garner et al., 2024; Robison & Nguyen, 2023). Therefore, this study also considered the combined impact of motivational messages in conjunction with signal words and alert symbols compared to when no message is displayed before completing a task.

The current study sought to explore the combined impact of various signal words/text highlight colors on performance related to vigilance-based tasks. Signal words are commonly used to denote hazards in the workplace, as they are attention-grabbing stimuli. Therefore, it was anticipated that using signal words and corresponding motivating messaging would increase accuracy on a basic vigilance task. Due to a lack of existing literature, no hypotheses were made for the directionality of crossover conditions (e.g., WARNING signal word with yellow text). However, it was expected that the following patterns would be seen for the primary condition combinations, that are the conditions where the signal word and text highlight color combinations are consistent with the ANSI (2002) Z535 recommendations (ANSI, 2002):

It was hypothesized that:

**H1 – Exp. 1**: Performance accuracy on the sustained attention task will improve within subjects when presented with a manipulation (signal word/text highlight color combination) compared to performance on a training block where no manipulation is present.

**H2 – Exp. 1**: Participants who receive the motivational message in conjunction with a signal word and alert symbol will perform significantly more accurately on the task compared to participants who do not receive a motivational message.

**H3 – Exp. 1:** Performance accuracy on the sustained attention task will decrease within subjects from the first half of the task compared to the second half of the task.

**H4 – Exp. 1**: Significant differences will be observed between the combined signal word/text highlight color conditions. Specifically, participants who receive the combined orange/WARNING prompt will have the highest accuracy on the sustained attention task, followed by participants who receive the combined yellow/CAUTION prompt, and those in the control condition will have the lowest accuracy out of any of the conditions.

Support for Hypotheses 1-4 would provide evidence that signal words and motivational messages are effective tools to mitigate attentional issues on tasks that require high vigilance levels. However, there are many potential factors that may contribute to vigilance decrements on sustained attention tasks, including various environmental and/or individual differences. Boredom is one individual difference that is relevant when examining the source of sustained attentional issues. In the following section, boredom's relationship with vigilance is reviewed in more detail.

**Sustained Attention and Boredom**

Within boredom research, distinctions have been made between the personality predisposition known as *trait boredom* or *boredom proneness*, and in-the-moment experiences of *state boredom* (Bench & Lench, 2019). Trait boredom constitutes a more stable personality construct that represents an individual's predisposition toward feeling bored (Farmer & Sundberg, 1986). On the other hand, state boredom describes an individual's fluctuations in experiencing boredom in a given situation, so it can vary based on environmental contexts (Fahlman et al., 2013). According to Mindlessness Theory, there should be a link between high experiences of boredom and accuracy decrements on sustained attention tasks. In one study, individuals with high trait boredom were found to have poor accuracy on a sustained attention task (Malkovsky et al., 2012). Another study found that trait boredom significantly predicted sustained attention performance (Hunter & Eastwood, 2018). There is also evidence that engagement in sustained attention tasks results in increased experiences of state boredom. One study found that performance on a basic sustained attention task paradigm known as the Sustained Attention to Response Task (SART; Robertson, 1997) increased self-reported state boredom (Hunter & Eastwood, 2018), and another study later replicated this finding in a different sample (Petranker & Eastwood, 2021).

Based on the existing empirical research, the relationship between performance on vigilance tasks and trait/state boredom were explored, as experiences of trait boredom (Hunter & Eastwood, 2018; Malkovsky et al., 2012) and state boredom (Hunter & Eastwood, 2018;

Petranker & Eastwood, 2021) have been linked to performance decrements on sustained attention tasks.

It was anticipated that the following patterns would be observed:

**H5 – Exp. 1:** Participants who receive the motivational message in conjunction with a signal word and alert symbol will have significantly lower state boredom compared to participants who do not receive a motivational message.

**H6 – Exp. 1**: Participants who receive any of the above-described signal word/text highlight color message combinations will have lower self-reported state boredom scores compared to participants in the control condition.

**H7 – Exp. 1**: Participants with higher trait boredom scores will have lower total score accuracy on the sustained attention task.

**H8 – Exp. 1**: Participants with higher state boredom scores will have lower total score accuracy on the sustained attention task.

Support for Hypotheses 5 and 6 would provide evidence that signal words/motivating messages are effective mitigation tools to increase attentional performance by reducing experiences of boredom. Evidence for Hypotheses 7 and 8 would support previous research that boredom is associated with reduced accuracy on sustained attention tasks. Specifically, it would suggest that both trait and state boredom are related to performance decrements on vigilance tasks (Hunter & Eastwood, 2018; Petranker & Eastwood, 2021). These findings would be insightful, but it is important to note that boredom is just one of many potential factors that may influence sustained attentional performance. An additional factor that may significantly impact performance outcomes on vigilance-based tasks is the presence or lack thereof of elements that introduce a sense of social motivation, including gamification and reward incentivization.

## Sustained Attention and Social Motivation

Many external factors have the potential to influence performance of tasks in both helpful and consequential ways. For example, social facilitation research demonstrates that performance of *simple* and *recognizable* tasks improves in the presence of others compared to when alone (Steinmetz & Pfattheicher, 2017; Tripplett, 1898; Zajonc, 1965). This phenomenon coincides with social inhibition, the experience of declining performance on *complex* tasks when others are watching (Bond & Titus, 1983; Guerin, 2010; Klehe et al., 2007; Latané, 1981; Steinmetz & Pfattheicher, 2017; Uziel, 2007; Zajonc & Sales, 1966). This experience has been documented even when participants are not being directly observed by others, such as in various research efforts where observers wore blindfolds and earplugs while a participant completed tasks (Markus, 1978; Platania & Moran, 2001, Schmitt et al., 1986; Steinmetz & Pfattheicher, 2017). Further, the model of evaluation apprehension (Cottrell, 1972) explains that the anticipation of future evaluation elicits similar responses to social facilitation effects, even if the participant is not being observed during the task itself. For example, in a study conducted by Ambach et al. (2019), the expectation that performance on a task would be evaluated later resulted in improved responses, even though no direct observers were present.

Another concept related to social facilitation is social motivation: the desire to strive for a particular goal or outcome based on the influence of social factors (Abrams & Hogg, 1990). Competition is one example of a socially motivating stimulus linked to increased attention and learning (DiMenichi & Tricomi, 2015). Competition is a complex stimulus, such that it can create goals and motivation to meet said goals, or it can create a negative sense of pressure and fear. The methodology underlying how competition is presented can influence perception of the stimulus both positively and negatively. For example, Reeve (2023) described motivationally constructive competition as competition that is informational and presented in a format that supports intrinsic motivation, compared to motivationally destructive competition that involves high pressure, ignored needs, and a focus on extrinsic motivation over intrinsic. One study found that in a physical effort-based task, competition resulted in significantly faster reaction times that was indicative of increased focus and attention on the task. However, the same study found that during a memory task, competition resulted in participants performing significantly worse than in the control condition (DiMenichi & Tricomi, 2015). As documented with social facilitation and social inhibition, the complexity of the task may have reinforced poor performance due to the fear of making a mistake and becoming embarrassed, either in that moment or later when performance was evaluated (Sanna & Shotland, 1990). It is also important to consider that competition can be presented in different methods, and it is possible that some formats may be more motivating than others. For example, some fitness companies have introduced a "gamification" model, where physical output is quantified, and users are challenged to beat others' scores as a motivator for having an effective workout. Other competition models are not focused on comparing performance against others but instead challenge users to meet a goal for the purpose of earning a reward. These types of strategies can result in prolonged attention on the task compared to if no stimulus was introduced. For example, there is evidence that when a financial reward is offered for performance, participant attention is maintained for a longer period compared to participants who complete the same task but where no financial reward is offered (Begleiter et al., 1983; Hömberg et al., 1981; Oken et al., 2010).

As sustaining attention is a cognitively demanding task that requires resources and processing over time (Davies & Parasuraman, 1982; Hancock & Warm, 1989; Hockey, 1997; Kahneman, 1973; Parasuraman et al., 1987; Wickens, 2002), it is necessary to maintain motivation to be successful (Oken et al., 2010). When looking at social motivation within the context of sustained attention, one study found that including competition and “points-based rewards” significantly reduced performance decrement over time (Robison & Nguyen, 2023, pp. 1256). This study’s results yielded evidence that sustained attention performance is based more on “willingness” as opposed to “capability” (Robison & Nguyen, 2023, pp. 1256). This information suggests that various incentivization types may help improve performance accuracy on sustained attention tasks by providing a motivator to keep the user engaged in the task at hand.

Based on the previous literature, the impact of various types of social incentives on vigilance-based task performance was examined in a separate experiment (Experiment Two). Specifically, the effects of gamification using a score ranking leaderboard and reward incentivization on sustained attention performance metrics were explored. Note that due to a lack of existing literature, no directionality hypotheses were made for whether the leaderboard or reward incentivization conditions would yield higher accuracy compared to the other.

However, the following was expected:

**H1 – Exp. 2:** Performance accuracy on the sustained attention task will improve within subjects when presented with a social motivation manipulation compared to performance on a training block where no manipulation is present.

**H2 – Exp. 2:** Participants who receive a social motivation manipulation will perform significantly more accurately on the task compared to participants who are in the control condition.

**H3 – Exp. 2:** Performance accuracy on the sustained attention task will decrease within subjects from the first half of the task compared to the second half of the task.

**H4 – Exp. 2:** Significant differences will be seen across the social motivation conditions. Specifically, participants will have the highest accuracy on the sustained attention task in the combined leaderboard/reward incentivization condition, where the most motivating stimuli are present, followed by participants in the standalone gamification leaderboard and reward incentivization conditions, and participants in the control condition will have the lowest accuracy out of any condition.

**H5 – Exp. 2:** Participants who receive the social motivation manipulation, as measured by the combined experimental groups, will have significantly lower state boredom compared to participants who do receive a social motivation manipulation.

**H6 – Exp. 2:** Participants who receive any of the three standalone social motivation manipulations will have lower self-reported state boredom scores compared to participants in the control condition.

**H7 – Exp. 2**: Participants with higher trait boredom scores will have lower total accuracy scores on the sustained attention task, regardless of the incentives offered.

**H8 – Exp. 2**: Participants with higher state boredom scores will have lower total accuracy scores on the sustained attention task, regardless of the incentives offered.

**Sustained Attention and Reaction Time**

Response accuracy and reaction time (RT) are often both considered when evaluating sustained attention and vigilance decrements (Yamashita et al., 2021). Quicker RTs are often seen prior to commission errors that occur when a participant incorrectly responds to a target. Slower RTs are often seen prior to errors of omission, when a participant fails to respond to a target (Allan et al., 2009; Yamashita et al., 2021). Previous research has suggested that faster RTs preceding commission errors depict instances of mindlessness and thoughts that are unrelated to the current task (Allan et al., 2009; Robertson et al., 1997; Smallwood et al., 2004; Yamashita et al., 2021), whereas slower RTs before omission errors are representative of a lack of engagement on the task (Allan et al., 2009; Yamashita et al., 2021). Therefore, both fast and slow reaction time speeds may represent dimensions of attention lapses (Yamashita et al., 2021). However, in a study conducted by Morais et al. (2024), participants with lower reaction times had significantly more correct detections and fewer incorrect and missed detections on a vigilance task. This finding may indicate that faster RTs could be related to stronger vigilance capabilities. For the current study, mean RT information was collected for each participant. As both fast and slow RTs can represent aspects of inattention, no directional hypotheses were made for the impact of RT. However, RT was explored as it relates to our other variables of interest. Specifically, exploratory analyses were conducted to assess the relationship between RT and task accuracy, trait boredom, and state boredom. Further, the impact of various signal word combinations and social motivators on RT were compared and measured.

**Study Overview**

The current project consisted of two experiments to assess various factors that may impact performance on vigilance-based tasks. Experiment One explored the effects of varying signal words on vigilance, boredom, and RT. Understanding the many potential factors and individual differences that impact sustained attention is critical in the design of user-centered technologies, as errors due to vigilance decrements have the potential to result in detrimental outcomes in many instances. Therefore, Experiment One examined the effects of signal word messaging on vigilance task accuracy as metrics of attentional performance. Additionally, the relationships between task performance accuracy, RT, and boredom were explored, as boredom is an individual difference often associated with reduced performance on vigilance tasks (Hunter

& Eastwood, 2018; Malkovsky et al., 2012; Robertson, 1997). Experiment Two assessed the effect of social motivation on vigilance, boredom, and RT. Specifically, the impact of various socially motivating elements was explored. Gamification's effect on sustained attention task accuracy, RT, and boredom compared to other competition formats was examined using a ranked point leaderboard. Additionally, the impact of social motivation in the form of reward incentivization was measured as a tool to motivate performance on vigilance tasks compared to other potential formats. The relationships between vigilance, RT, and boredom were also explored.

## Experiment One

Experiment One sought to examine differences in accuracy and RT on a basic vigilance task when various signal word/text highlight color message combinations preceded the task. An additional component of this study involved exploring the relationships between trait boredom, state boredom, RT, and accuracy on a basic vigilance task. Together, the overarching goal of Experiment One was to inform the design of user-centered protocols for roles that require high sustained attention. Specifically, understanding the effectiveness of pairing various signal words with motivating messages to capture attention and improve performance could be invaluable in improving efficiency and reducing errors across many domains.

### Method

#### *Participants*

Data collection took place remotely from April 2024 through September 2024. All participants were undergraduate students enrolled in an Introduction to Psychology course at North Carolina State University. Participants were recruited through this course, as a syllabus requirement was to select between participating in research or completing an alternative written assignment of equal effort. All participants were required to have access to a computer with a keyboard and internet connection to complete the study. This study's procedures were approved by an institutional review before data collection began. Prior to beginning the study, participants read through and acknowledged informed consent documentation.

A power analysis was conducted using G* Power 3 software (Faul et al., 2007) to determine the necessary sample size. The analysis was conducted based on the Kruskal-Wallis H test, the statistical test that required the largest sample size among the analysis plan, to ensure adequate power for all tests. However, the software does not provide a built-in option for the Kruskal-Wallis H test, so the analysis was conducted using a one-way ANOVA model as an approximation. With an effect size of .30 and power of .80, at least 161 participants were required. Ultimately, 190 participants were recruited for the study, but 19 participants were excluded from analyses. Entries were excluded due to missing data ($n = 6$), not passing the adapted color vision screener ($n = 6$), self-reporting that the participant had night mode and/or reverse contrast enabled on the computer during the study ($n = 3$), data that suggested a lack of engagement with the task (never pressed the space bar during the 50-minute full duration of the task) ($n = 1$), or for extreme outlying scores on the Mackworth Clock Test (>3 *SD*s below the mean after removing all other data excluded from analyses and consistent with the most extreme outliers based on the Interquartile Range (IQR)) ($n = 3$). Note that only the three most extreme outliers were excluded from analyses because they were indicative that the participant did not appropriately complete the task. However, less extreme outliers (datapoints that were within three *SD*s of the mean but were identified as outliers based on the IQR) remained in the analyses, as they appeared to be true outliers based on performance and not based on data noise. Therefore, the final sample consisted of 171 participants (Group 1: $n = 23$, Group 2: $n = 25$, Group 3: $n =$

34, Group 4: *n* = 25, Group 5: *n* = 18, Group 6: *n* = 25, Group 7: *n* = 21). Eighty-nine participants identified as male, 80 participants identified as female, and 2 participants preferred not to answer. Ages ranged from 18 to 31 years (*M* = 19.49, *SD* = 1.71), and participants identified as the following ethnicities: 5.8% Black/African American, 11.1% Asian American/Pacific Islander, 66.7% White/Caucasian, 5.8% Hispanic/Latino, 3.5% multiple ethnicities, and 7.0% preferred not to respond.

### *Design*

This study used a 7x2 mixed factorial design to explore the combined effect of signal word and text highlight color on vigilance task accuracy, RT, and boredom. The first independent variable, combined signal word/text highlight color, was manipulated between-subjects across seven signal word/text highlight color groups. The second independent variable, time of measurement, was manipulated within-subjects before the manipulation was introduced (directly after the ten-minute training block) compared to after the manipulation was introduced (ten minutes after the manipulation was presented). Task accuracy, RT, and state boredom were measured as dependent variables with comparisons made between the first and second half of the vigilance task. Additionally, trait boredom was measured as a covariate.

### *Independent Variable*

*Combined Signal Word/Text Highlight Color Prompt.* A motivational message was included in all experimental conditions as an extrinsic motivator intended to promote sustained attention task accuracy. There were six versions of the same message used by McGough & Mayhorn (2022, 2023) with varying signal word and text highlight color combinations. Of note, standalone text highlight color conditions were not included, as they lacked meaning for the purposes of this study. Additionally, there was one control condition, where no message or alert symbol was presented, and participants instead began immediately with the vigilance task. The conditions were as follows: control, orange/WARNING, yellow/WARNING, no color/WARNING, orange/CAUTION, yellow/CAUTION, no color/CAUTION. As shown in Figure 1, the prompts read:

> "**WARNING**: If you do not complete the following task with at least **95%** accuracy, you will be required to re-start the activity."
>
> "**CAUTION**: If you do not complete the following task with at least **95%** accuracy, " you will be required to re-start the activity."

Additionally, each prompt's signal word was highlighted in the color of the respective condition, and the prompts included an alert symbol based on the ANSI (2002) Z535 standards (ANSI, 2002).

**Figure 1**

*Signal Word/Text Highlight Color Variations.*


*Note.* The control group, who received no signal word/text highlight color/exclamation point within a triangle alert symbol, is not pictured because participants in this group viewed no image prior to beginning the task.

***Dependent Measures***

*Mackworth Clock Test (Mackworth, 1948).* This test was originally developed as a physical device resembling a clock that was used to assess a person's vigilance. Norman Mackworth designed the tool to assess World War II radar operators' abilities to focus on a task and attend to targets over time. His findings showed that when completing the task, decrements in vigilance were seen within thirty minutes of the task beginning. Since then, computer-based paradigms of the task have been developed (Lichstein, Riedel, & Richman, 2000). During the task, participants were instructed to watch a clock's hand move around a clock face. When the hand of the clock moved a farther distance than normal, they were asked to press the spacebar on their keyboard. Individual trial accuracy feedback was given every time the participant clicked the space bar. A green light appeared if the trial was correct, and a red light appeared if the trial was incorrect. Further, if the participant failed to press the spacebar on a trial where the hand moved more than normal, a red light appeared to inform the participant they missed that trial. Participants had one second to respond each time the clock hand moved its place. In the current study, participants first completed a training period of ten minutes and then completed a 50-minute trial. The rate of clock hand "jumps" was slightly more than half a percent (0.667%), consistent with Mackworth's original test design. In the ten-minute training period, there were approximately 600 total trials and 4 jump trials. In the 50-minute full task, there were approximately 3,000 total trials and 20 jump trials. However, these numbers varied by participant based on task behavior and response type, since the accuracy feedback provided after correct detections, missed detections, and incorrect detections increased the timing of these individual trials. RT (measured in milliseconds), total number of trials, correct detections, missed detections, and incorrect detections were recorded for each participant. Additionally, two accuracy proportions were calculated. The first calculation was a proportion of the total number of “correct” trials, with “correct” meaning that the participant responded appropriately to that given trial (e.g., correctly pressed the space bar on jump trials and did not press the spacebar on non-jump trials). This number was referred to as the Correct Response ratio. The second calculation was a proportion of accurate responses to correct detection trials. This number was referred to as the Correct Detection ratio. All accuracy analyses were conducted using each proportion individually to get a complete sense of accuracy across the full task.

$$Correct\ Response\ Ratio = (Total\ \#\ of\ Trials\ -\ Missed\ Detections\ - \\ Incorrect\ Detections)\ \div\ Total\ \#\ of\ Trials$$

$$Correct\ Detection\ Ratio =\ (Total\ \#\ of\ Correct\ Detections\ \div\ Total\ \#\ of\ Jump\ Trials)$$

*Boredom Proneness Scale* (BPS; Farmer & Sundberg, 1986). The Boredom Proneness Scale measured trait boredom, defined as an individual’s predisposition towards experiencing boredom. In other words, the scale provided information about how easily someone becomes bored. It consisted of 28 questions scored along a 7-point Likert scale (Highly Disagree (1) to Highly Agree (7)). This scoring format was a revision of the scale's original True/False design (Vodanovich & Watt, 2016). A higher total score indicated a higher trait boredom level, and a lower total score indicated a lower trait boredom level. Example items from the scale are as follows: “I find it easy to entertain myself” and “It takes a lot of change and variety to keep me really happy.” The current sample’s Cronbach alpha indicated good reliability (α = .819).

*Multidimensional State Boredom Scale* (MSBS; Fahlman et al., 2013). The Multidimensional State Boredom Scale measured state boredom, an “in the moment” experience of boredom. It consisted of 29 questions across five subscales (Disengagement, High Arousal, Inattention, Low Arousal, and Time Perception). Each question was scored using a 7-point Likert scale (Strongly Disagree (1) to Strongly Agree (7)). A higher total score represented a greater experience of boredom, and a lower score indicated less boredom. Example items from the questionnaire are as follows: “Everything seems repetitive and routine to me” and “I feel cut off from the rest of the world.” Internal consistency for the current sample indicated excellent reliability ($\alpha = .939$).

*Ishihara Test* (Ishihara, 1917). This test was designed to screen for color vision deficiencies, and the full test consists of 38 plates that participants view. In a reduced version administered for the current study, participants viewed ten images and were asked to report what number they saw. For this study, data from participants who answered more than two questions incorrectly were excluded from analyses ($n = 6$), as incorrect answers may suggest that a person experiences color vision deficiencies, and experiences of color vision deficiencies may have influenced how a participant responded in the study. Of note, this test was not considered to be diagnostic for the current study. Therefore, answering two or more questions wrong was used as a threshold criterion to be consistent across all participants, but scores were not representative or suggestive of any type of diagnosis. Additionally, even if a participant scored more than two questions incorrectly, they were still permitted to participate in the study and earned full research credit.

*Demographic Questionnaire.* A demographic questionnaire was administered at the end of the study to assess the representativeness of the current sample. Specifically, participants provided information about age, gender, ethnicity, color vision deficiencies, and information about their computer displays. Only participants 18 and older were permitted to complete the study and earn credit for their participation.

***Procedure***

Upon signing up for the study, all participants received a link for PsyToolkit, an online repository for psychological experiments and data collection (Stoet 2010, 2017). First, participants read through and virtually acknowledged the informed consent documentation. If they agreed to the informed consent, the participant was directed to complete the study.

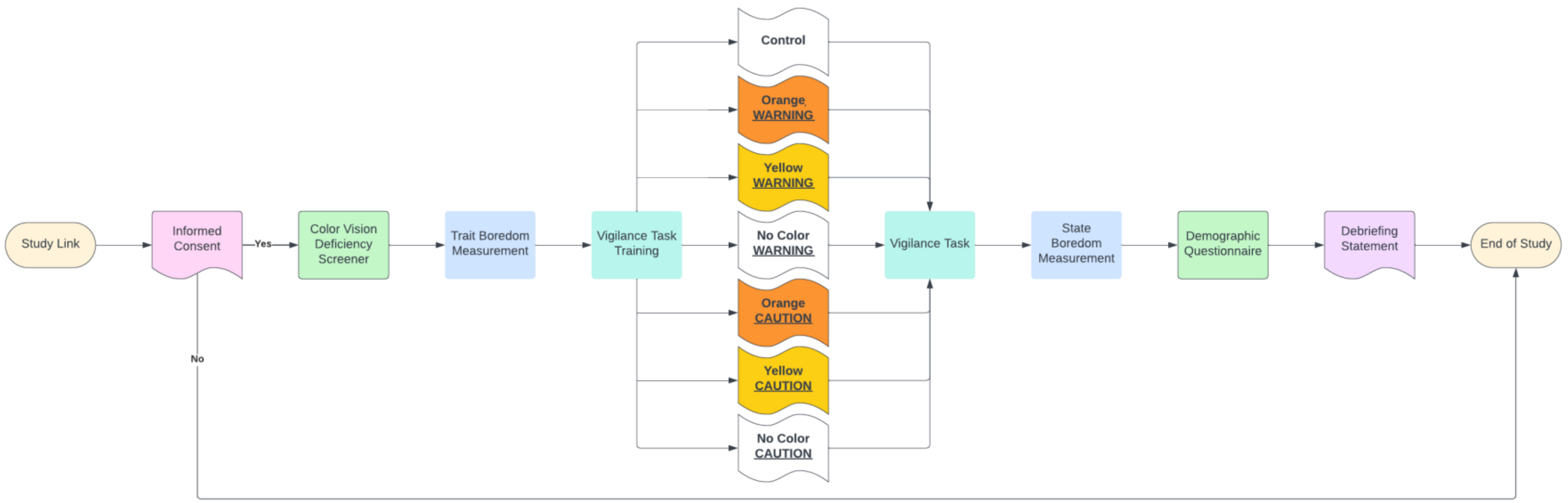
On the following screen, participants were instructed to enter their SONA ID, a unique identification number used to assign research credit. SONA Systems is an online participant recruitment and study management platform (SONA Systems, 2024) that participants used to enroll in research studies. After entering their unique ID, participants were asked to confirm that their screen's display brightness was at full brightness and that they did not have night shift and/or reversed contrast enabled on their computer. Participants were also reminded at this time that if they wore glasses or contacts, they should wear them for the duration of the study. Then, participants completed a condensed version of the Ishihara Test, a screener for color vision deficiencies (Ishihara, 1917), followed by the Boredom Proneness Scale (Farmer & Sundberg, 1986).

After completing the Boredom Proneness Scale (Farmer & Sundberg, 1986), all participants viewed instructions for the Mackworth Clock Test (Mackworth, 1948) and subsequently completed a ten-minute training block to become familiar with the task. After, they were notified that the training block had ended, and they were now beginning the actual task. The instructions for the task were shown on the screen once again. Then, based on their assigned study condition, participants viewed a language prompt with one of the following combinations: WARNING/CAUTION and orange/yellow/no color, or a control condition (no language/no color). The message read "[SIGNAL WORD]: If you do not complete the following task with at least **95%** accuracy, you will be required to re-start the activity." All participants then completed the 50-minute task.

At the end of the vigilance task, participants completed the Multidimensional State Boredom Scale (Fahlman et al., 2013) and a short Demographics Questionnaire. Participants were also asked if they experienced any type of color vision deficiency, and they were asked to provide basic information about their display screen size. After the study, all participants viewed a debriefing statement with an option available to download the document for their records. In total, the study took approximately 90 minutes to complete. For this experiment, very weak deception, as approved in advance of data collection by an institutional review board, was used such that participants did not actually have to restart the study if they scored below a certain threshold. Rather, this message and its corresponding signal word/text highlight color variations were included to explore their impact on other performance metrics, such as accuracy, RT, and boredom. Figure 2 visually depicts the procedural flow for Experiment One.

**Figure 2**

*Procedural Flow for Experiment One*



***Statistical Analyses***

All data was evaluated for normality by examining the skewness and kurtosis levels (>2), visually inspecting the histograms for each variable's data, and examining the Shapiro Wilks and Kolmogorov-Smirnov Tests. The trait boredom (BPS) and state boredom (MSBS) variables were normally distributed. However, the accuracy variables for the Mackworth Clock Test training, Correct Response ratio, and Correct Detection ratio were negatively skewed. Additionally, the RT variable had non-normal distributions. Therefore, nonparametric analyses were performed for all analyses involving the vigilance task accuracy and RT variable, and any analyses that focused solely on boredom variables were parametric. Table 1 and 2 outline the descriptive statistics for the sample.

**Table 1**

*Descriptive Statistics – Mackworth Clock Test Accuracy, Trait Boredom, State Boredom*

| | Total | | | | Control | | | | Experimental | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *N* | *M* | *SD* | *Mdn* | *N* | *M* | *SD* | *Mdn* | *N* | *M* | *SD* | *Mdn* |
| Training – Correct Responses | 171 | 0.971 | 0.075 | 0.992 | 23 | 0.990 | 0.007 | 0.992 | 148 | 0.968 | 0.080 | 0.992 |
| Correct Responses | 171 | 0.996 | 0.008 | 0.998 | 23 | 0.995 | 0.003 | 0.995 | 148 | 0.996 | 0.009 | 0.998 |
| Correct Detections | 171 | 0.740 | 0.292 | 0.850 | 23 | 0.465 | 0.363 | 0.552 | 148 | 0.782 | 0.256 | 0.862 |
| Incorrect Detections | 171 | 0.002 | 0.007 | 0.001 | 23 | 0.002 | 0.001 | 0.002 | 148 | 0.003 | 0.007 | 0.001 |
| Missed Detections | 171 | 0.260 | 0.292 | 0.150 | 23 | 0.535 | 0.363 | 0.448 | 148 | 0.217 | 0.255 | 0.138 |
| Trait Boredom | 171 | 102.164 | 16.804 | 102.000 | 23 | 99.913 | 15.409 | 101.000 | 148 | 102.514 | 17.032 | 103.000 |
| State Boredom | 171 | 119.813 | 29.531 | 118.000 | 23 | 126.696 | 23.686 | 131.000 | 148 | 118.743 | 30.265 | 116.500 |

*Note.* Correct Response ratios were calculated by subtracting missed detections and incorrect detections from the total number of trials and dividing that score by the total number of trials. Correct Detection and Missed Detection ratios were calculated out of the total number of jump (target) trials. The Incorrect Detection ratio was calculated out of the total number of trials.

**Table 2**

*Descriptive Statistics – Correct Detections RT*

| | *N* | *M* | *SD* | *Mdn* |
|---|---|---|---|---|
| Total | 161 | 607.62 | 73.42 | 600.62 |
| Control | 19 | 650.56 | 74.73 | 658.06 |
| Experimental | 142 | 601.87 | 71.57 | 593.20 |

*Note.* RT was calculated on trials where participants successfully pressed the space bar on Correct Detection targets. Eight participants, while having engaged with the task, failed to successfully detect any Correct Detection trials, so there was no RT data to represent Correct Detection RT for these participants. Two additional participants were excluded due to outlying scores (>3 *SD*s outside the mean RT score and consistent with the most extreme outliers based on the IQR).

## Results

***Mackworth Vigilance Test Task Performance.*** Vigilance task performance was captured by examining accuracy data for the training block and examining accuracy and RT data for the full task. Task accuracy was measured in two formats: the Correct Response ratio and the Correct Detection ratio. See *Measures* for the full equation information. Exploratory RT analyses were conducted by taking the mean reaction time for correct detections (Correct Go/Jump trials).

A Pairwise Mann Whitney U test was used to examine gender differences in vigilance task accuracy variables (Correct Response ratio and Correct Detection ratio) and the vigilance task training variable (Correct Response ratio). The gender variable had several response options, but the participant responses ultimately were divided between male, female, and two participants who preferred not to answer. The “preferred not to answer” option was not considered a separate gender category, so the analysis was performed using only the male and female groups. Results of the analysis indicated a significant difference in Training Accuracy (Training Correct Response ratio) on the Mackworth Clock Test between males (*Mdn* = 0.990) and females (*Mdn* = 0.996), $U = 2804.50$, $z = -2.38$, $p = .017$, where females performed significantly more accurately than males. No significant gender differences were found in vigilance task accuracy for the Correct Response ratio or Correct Detection ratio. An additional Mann Whitney U Test was conducted to examine gender differences in RT. Results indicated a significant RT difference between males (*Mdn* = 578.00) and females (*Mdn* = 616.15), $U = 2354.00$, $z = -2.77$, $p = .006$, where males had significantly faster RTs on Correct Detection trials than females.

The “preferred not to answer” option on the Ethnicity Demographic question was also not considered a separate category, so ethnicity difference analyses were performed excluding this option. No significant ethnicity differences were found in the vigilance task Correct Response ratio, Training Accuracy (Training Correct Response ratio), or RT variables. However, significant ethnicity differences were seen in the vigilance task Correct Detection ratio, $\chi^2(4) = 11.45$, $p = .022$. Post hoc comparisons using Dunn's Test with Bonferroni Corrections showed significant differences between participants who identified as Asian American/Pacific Islander (*Mdn* = 0.609) and participants who identified as Hispanic/Latino (*Mdn* = 0.971), adjusted $p = .019$.

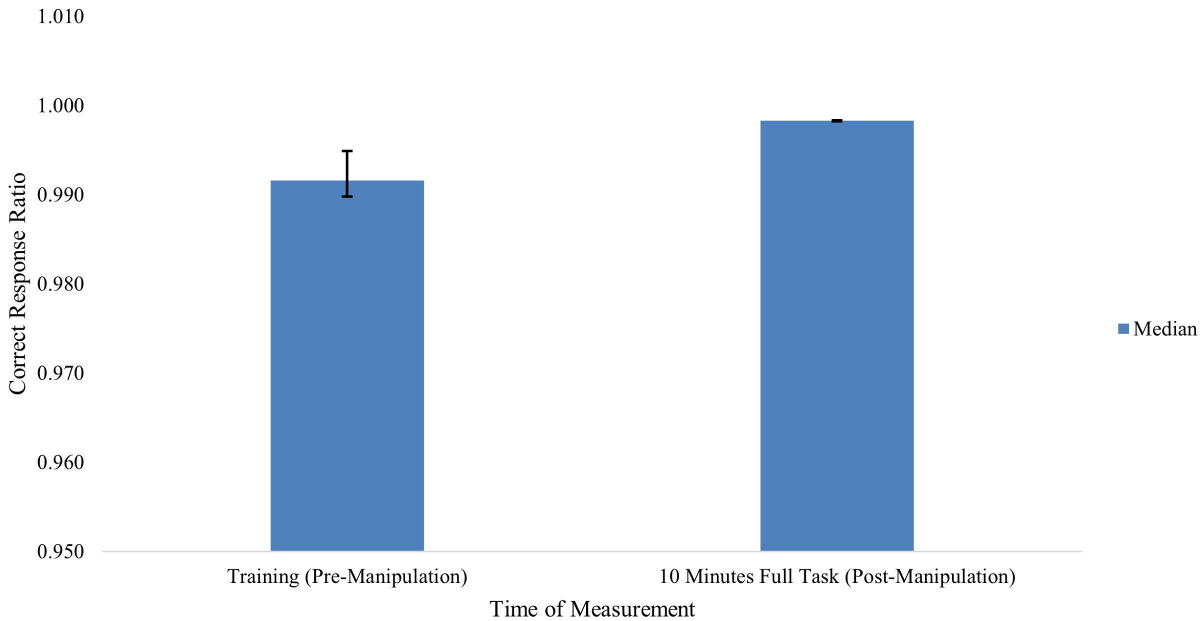
*Training.* All participants completed an identical ten-minute version of the vigilance task, referred to as the "training block", prior to completing the full 50-minute task. This training period was administered to serve as a within-subjects performance comparison before versus after the manipulations were presented. A Sign Test was used to compare performance accuracy on the vigilance task between the ten-minute pre-manipulation training block and the first ten minutes of the task post-manipulation (Correct Response ratio). Accuracy during the first ten minutes of the actual task was calculated in analysis by using the same number of trials as the participant completed in the training block to compare performance. Of note, for all analyses involving the training block and the ten-minute full task comparison, results were calculated using only the Correct Response ratio and not the Correct Detection ratio, as the number of jump trials in each ten-minute window contained extremely limited data points (~4 jump trials). Therefore, RT analyses were not conducted for the training block, as RT was based on correct detections (See *Full Vigilance Task – Reaction Time*). Further, as these analyses sought to compare differences based on the manipulation, participants in the control condition were excluded.

A Sign Test was conducted to compare the Correct Response ratio on the vigilance task before and after the manipulation was introduced, as the data's distribution violated the

Wilcoxon's Sign-Rank test's assumption that the distribution of the differences between the two groups are symmetrical (H1 – Exp. 1). For this analysis, the Correct Response ratio in the ten-minute training block was compared to the Correct Response ratio in the first ten minutes of the actual task, after the signal word/text highlight color manipulation was introduced. As this analysis sought to compare Correct Response ratio differences based on the manipulation, participants in the control condition were excluded. Results indicated a significant difference, with 26 negative differences, 97 positive differences, and 25 ties, $z = -6.31$, $p < .001$. This finding suggests that participants performed significantly more accurately (as measured by overall correct responses) on the vigilance task after receiving the signal word/text highlight color message (*Mdn* = 0.998) compared to in the training block (*Mdn* = 0.992). This finding provides support for Hypothesis 1 – Exp. 1, which stated that "Performance accuracy on the sustained attention task will improve within subjects when presented with a manipulation (signal word/color combination) compared to performance on a training block where no manipulation is present." Figure 3 depicts the within-subjects Correct Response ratio differences between the training block and the first 10 minutes of the vigilance task.

**Figure 3**

*Correct Response Ratio Differences Between Training and First Ten Minutes of Full Task. Y-axis begins at 0.950 to highlight visual differences.*



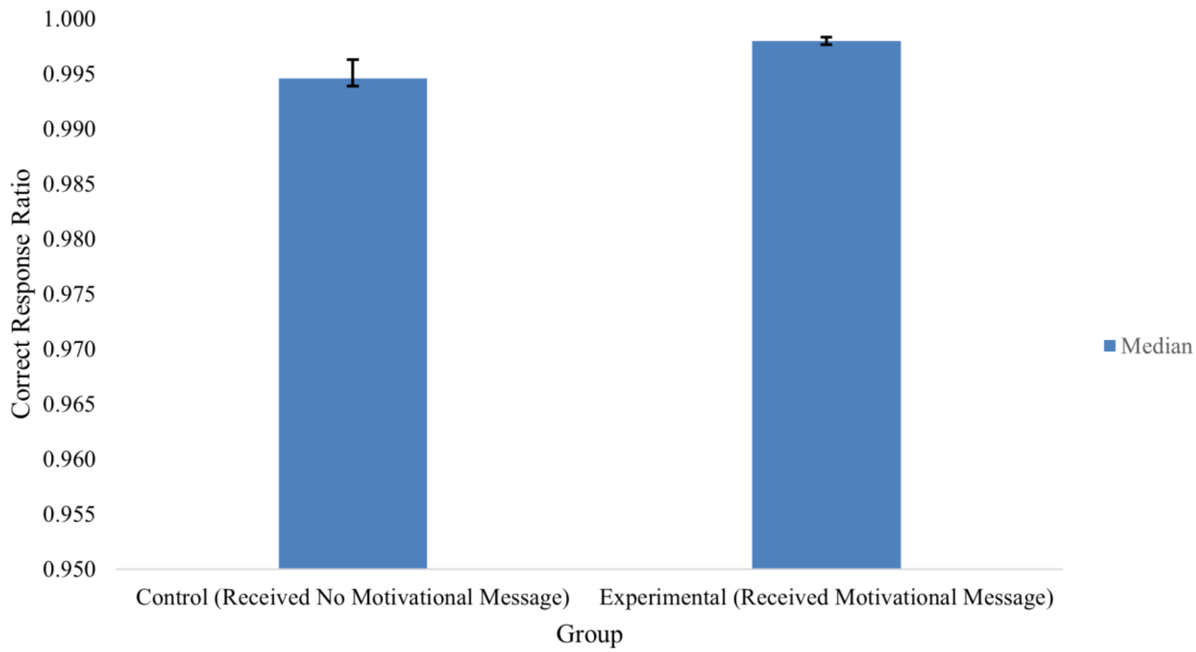

*Note.* Error bars indicate bootstrapped 95% confidence intervals (CI), calculated from 5,000 samples.

*Full Vigilance Task - Correct Responses.* A Mann-Whitney U Test was performed to compare the full task performance accuracy, as measured by the Correct Response ratio, of participants who received the motivational message in conjunction with a signal word and alert symbol compared to participants who received no message prior to completing the task. For this analysis, all participants across the six signal word/text highlight color groups that received the same motivational message were combined into one experimental group to compare with the control group, who received no message. A Sign Test was used to compare performance accuracy, as measured by the Correct Response ratio, on the full vigilance task between the first half of the 50-minute task and the second half of the 50-minute task. Additionally, a Kruskal-Wallis H Test was performed to assess the effects of the combined factor of signal word and text highlight color on the Correct Response ratio.

*Experimental vs. Control.* A Mann-Whitney U test was performed to compare the vigilance task accuracy of participants who received the motivational message in conjunction with a signal word and alert symbol compared to participants who received no message prior to completing the task (H2 – Exp 1: Correct Responses). For this analysis, the group that received the motivational message was the combined six experimental groups from the study, who all received the same motivational message but had varying signal word/text highlight colors. When accuracy was measured using the Correct Response ratio, results indicated significant differences between the control group (*Mdn* = 0.995, Mean Rank = 55.04) and groups who received the message (*Mdn* = 0.998, Mean Rank = 90.81), $U = 990.00$, $z = -3.22$, $p = .001$, where participants who received the combined motivational message, signal word, and alert symbol performed significantly more accurately, as measured by the Correct Response ratio, than participants in the control group. This finding provides support for Hypothesis 2 – Exp 1: Correct Responses, which stated that "Participants who received the motivational prompt in conjunction with the signal word and alert symbol will perform significantly more accurately on the task compared to participants who do not receive a motivational prompt." Figure 4 depicts the Correct Response ratio differences between participants who received no message prior to the task and those who received the motivational message, signal word, and alert symbol.

**Figure 4**

*Correct Response Ratio Differences Between Control and Experimental Groups. Y-axis begins at 0.950 to highlight visual differences.*




*Note.* Error bars indicate bootstrapped 95% confidence intervals (CI), calculated from 5,000 samples.

*First Half vs. Second Half.* A Sign Test was conducted to compare performance accuracy, as measured by the Correct Response ratio, on the vigilance task between the first half of the task and the second half of the task (H3 – Exp. 1: Correct Responses), as the data's distribution violated the Wilcoxon's Sign-Rank test's assumption that the distribution of the differences between the two groups are symmetrical. Results showed no significant Correct Response ratio differences between the first ($Mdn$ = 0.998) and second halves ($Mdn$ = 0.998) of the task, with 62 negative differences, 79 positive differences, and 30 ties, $z$ = -1.35, $p$ = .178, contrary to Hypothesis 3 – Exp. 1: Correct Responses, which stated that "Performance accuracy on the sustained attention task will decrease within subjects from the first half of the task compared to the second half of the task."

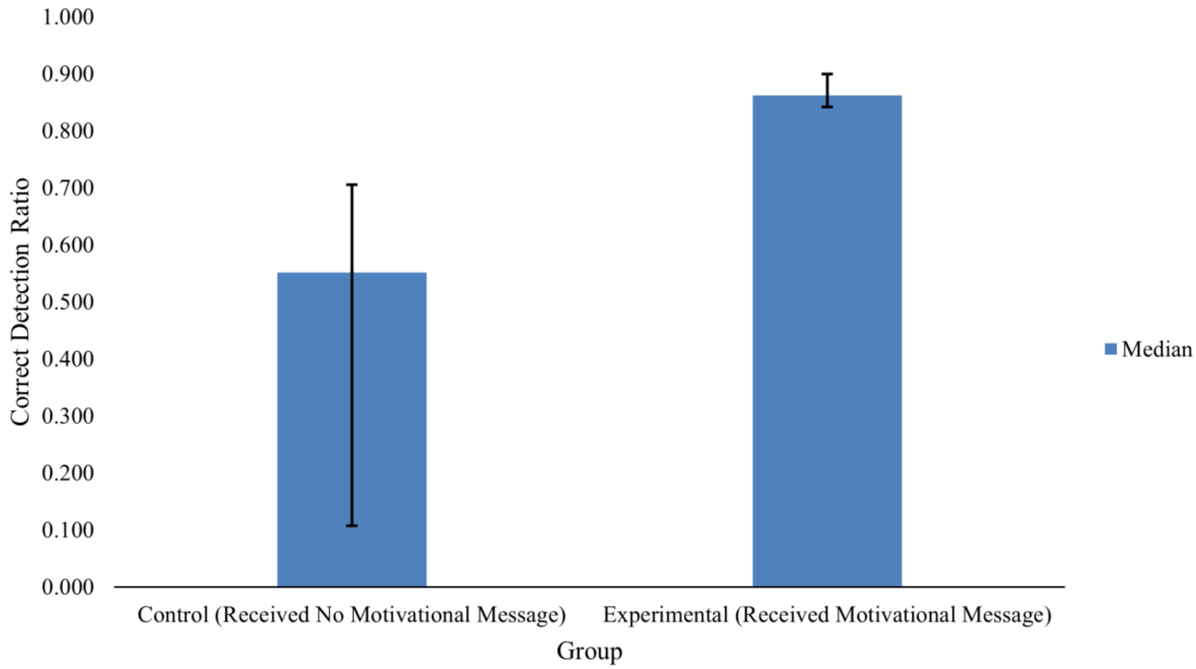
*Signal Word/Text Highlight Color Comparison.* A Kruskal-Wallis H test was performed to assess the effects of the combined factor of signal word and text highlight color on the vigilance task Correct Response ratio (H4 – Exp. 1: Correct Responses). The analysis showed no significant differences in the Correct Response ratio between the groups, $\chi^2(6) = 11.89$, $p = .064$, contrary to the hypothesis. The median and mean rank Correct Response ratio scores for each group was: Group 1 (control, *Mdn* = 0.995, Mean Rank = 55.04), Group 2 (orange/WARNING, *Mdn* = 0.998, Mean Rank = 89.90), Group 3 (yellow/WARNING, *Mdn* = 0.998, Mean Rank = 95.18), Group 4 (no color/WARNING, *Mdn* = 0.998, Mean Rank = 94.64), Group 5 (orange/CAUTION, *Mdn* = 0.997, Mean Rank = 79.83), Group 6 (yellow/CAUTION, *Mdn* = 0.997, Mean Rank = 87.44), Group 7 (no color/CAUTION, *Mdn* = 0.998, Mean Rank = 93.69). This finding did not support Hypothesis 4 – Exp. 1: Correct Responses, which stated that "Significant differences will be observed between the combined signal word/text highlight color conditions. Specifically, participants who receive the combined orange/WARNING prompt will have the highest accuracy on the sustained attention task, followed by participants who received the combined yellow/CAUTION prompt, and those in the control condition will have the lowest accuracy out of any of the conditions."

*Full Vigilance Task - Correct Detections.* A Mann-Whitney U Test was performed to compare the full task performance accuracy, as measured by the Correct Detection ratio, of participants who received the motivational message in conjunction with a signal word and alert symbol compared to participants who received no message prior to completing the task. A Sign Test was used to compare performance accuracy, as measured by the Correct Detection ratio, on the full vigilance task between the first half of the 50-minute task and the second half of the 50-minute task. Additionally, a Kruskal-Wallis H Test was performed to assess the effects of the combined factor of signal word and text highlight color on the Correct Detection ratio.

*Experimental vs. Control.* A Mann-Whitney U test was performed to compare the vigilance task accuracy of participants who received the combined motivational message, signal word, and alert symbol compared to participants in the control group, who received no message prior to completing the task (H2 – Exp. 1: Correct Detections). For this analysis, the group that received the motivational message was the combined six experimental groups from the study, who all received the same motivational message but had varying signal word/text highlight colors. When accuracy was measured using the Correct Detection ratio, the results revealed significant differences between the control group (*Mdn* = 0.552, Mean Rank = 46.61) and the experimental groups (*Mdn* = 0.862, Mean Rank = 92.12), such that participants who received the motivational message, signal word, and alert symbol had significantly more correct detections than participants in the control group, $U = 796.00$, $z = -4.11$, $p < .001$. This finding provides additional support for Hypothesis 2 – Exp. 1: Correct Detections, which stated that "Participants who received the motivational message in conjunction with a signal word and alert symbol will perform significantly more accurately on the task compared to participants who do not receive a motivational message." Figure 5 depicts the Correct Detection ratio differences between participants who received no motivational message prior to the task and those who received the motivational message.

**Figure 5**

*Correct Detection Ratio Differences Between Control and Experimental Groups*




*Note.* Error bars indicate bootstrapped 95% confidence intervals (CI), calculated from 5,000 samples.

*First Half vs. Second Half.* A Sign Test was used to compare performance accuracy, as measured by the Correct Detection ratio, on the full vigilance task between the first half of the 50-minute task and the second half of the 50-minute task. No significant differences were observed in the Correct Detection ratios between the first half (*Mdn* = 0.833) and second half (*Mdn* = 0.857) of the task, with 68 negative differences, 66 positive differences, and 37 ties, $z = -0.09$, $p = .931$. This finding is consistent with the same analysis conducted with the Correct Response ratio as the outcome variable but was contrary to Hypothesis 3 – Exp. 1: Correct Detections, which stated that "Performance accuracy on the sustained attention task will decrease within subjects from the first half of the task compared to the second half of the task."

*Signal Word/Text Highlight Color Comparison.* A Kruskal-Wallis H test was conducted to assess the effects of the combined factor of signal word and text highlight color on the Correct

Detection ratio (H4 – Exp. 1: Correct Detections). Results indicated a significant difference between groups, $\chi^2(6) = 18.54$, $p = .005$. Post hoc comparisons were made using Dunn's Test with Bonferroni Corrections. Results showed significant differences between several signal word/text highlight color groups. The median and mean rank scores for each group were as follows: Group 1 (control, *Mdn* = 0.552, Mean Rank = 46.61), Group 2 (orange/WARNING, *Mdn* = 0.885, Mean Rank = 95.16), Group 3 (yellow/WARNING, *Mdn* = 0.866, Mean Rank = 95.09), Group 4 (no color/WARNING, *Mdn* = 0.895, Mean Rank = 95.98), Group 5 (orange/CAUTION, *Mdn* = 0.812, Mean Rank = 81.19), Group 6 (yellow/CAUTION, *Mdn* = 0.895, Mean Rank = 94.32), Group 7 (no color/CAUTION, *Mdn* = 0.850, Mean Rank = 85.86). Results of the post hoc comparison indicate that Group 1 (control) performed significantly less accurately (as measured by the Correct Detection ratio) than Group 2 (orange/WARNING), Group 3 (yellow/WARNING), Group 4 (no color/WARNING), and Group 6 (yellow/CAUTION). Therefore, unlike when measuring by the Correct Response ratio, partial support was found for Hypothesis 4 – Exp. 1: Correct Detections when accuracy was measured by the Correct Detection ratio. Hypothesis 4 – Exp 1: Correct Detections stated that "Significant differences will be observed between the combined signal word/text highlight color conditions. Specifically, participants who receive the combined orange/WARNING prompt will have the highest accuracy on the sustained attention task, followed by participants who received the combined yellow/CAUTION prompt, and those in the control condition will have the lowest accuracy out of any of the conditions." As predicted by the hypothesis, the control group had the lowest Correct Detection ratio score across any of the signal word/text highlight color conditions and was significantly less accurate than several of the other conditions. However, no other groups were significantly different in accuracy compared to others. Table 3 outlines the pairwise comparisons.

**Table 3**

*Dunn's Multiple Comparison Test - Correct Detection Ratio*

| Comparison | Group | Test Statistic | Std. Error | Std. Test Statistic | *p* | Adj. *p*[a] |
|---|---|---|---|---|---|---|
| 1-5 | Control vs. Orange/CAUTION | -34.586 | 15.543 | -2.225 | .026 | .547 |
| 1-7 | Control vs. No Color/CAUTION | -39.248 | 14.907 | -2.633 | .008 | .178 |
| 1-6 | Control vs. Yellow/CAUTION | -47.711 | 14.270 | -3.343 | < .001 | .017* |
| 1-3 | Control VS. Yellow/WARNING | -48.480 | 13.334 | -3.636 | < .001 | .006** |
| 1-2 | Control vs. Orange/WARNING | -48.551 | 14.270 | -3.402 | < .001 | .014* |
| 1-4 | Control vs. No Color/WARNING | -49.371 | 14.270 | -3.460 | < .001 | .011* |
| 5-7 | Orange/CAUTION vs. No Color/CAUTION | -4.663 | 15.864 | -0.294 | .769 | 1.000 |
| 5-6 | Orange/CAUTION vs. Yellow/CAUTION | -13.126 | 15.267 | -0.860 | .390 | 1.000 |
| 5-3 | Orange/CAUTION vs. Yellow/WARNING | 13.894 | 14.397 | 0.965 | .335 | 1.000 |
| 5-2 | Orange/CAUTION vs. Orange/WARNING | 13.966 | 15.267 | 0.915 | .360 | 1.000 |
| 5-4 | Orange/CAUTION vs. No Color/WARNING | 14.786 | 15.267 | 0.968 | .333 | 1.000 |
| 7-6 | No Color/CAUTION vs. Yellow/CAUTION | 8.463 | 14.620 | 0.579 | .563 | 1.000 |
| 7-3 | No Color/CAUTION vs. Yellow/WARNING | 9.231 | 13.708 | 0.673 | .501 | 1.000 |
| 7-2 | No Color/CAUTION vs. Orange/WARNING | 9.303 | 14.620 | 0.636 | .525 | 1.000 |
| 7-4 | No Color/CAUTION vs. No Color/WARNING | 10.123 | 14.620 | 0.692 | .489 | 1.000 |
| 6-3 | Yellow/CAUTION vs. Yellow/WARNING | 0.768 | 13.012 | 0.059 | .953 | 1.000 |
| 6-2 | Yellow/CAUTION vs. Orange/WARNNING | 0.840 | 13.970 | 0.060 | .952 | 1.000 |
| 6-4 | Yellow/CAUTION vs. No Color/WARNING | 1.660 | 13.970 | 0.119 | .905 | 1.000 |
| 3-2 | Yellow/WARNING vs. Orange/WARNING | 0.072 | 13.012 | 0.006 | .996 | 1.000 |
| 3-4 | Yellow/WARNING vs. No Color/WARNING | -0.892 | 13.012 | -0.069 | .945 | 1.000 |
| 2-4 | Orange/WARNING VS. No Color/WARNING | -0.820 | 13.970 | -0.059 | .953 | 1.000 |

a. Bonferroni correction, $*p < .05$, $**p < .01$

*Full Vigilance Task - Reaction Times.* All analyses involving RT data were exploratory. A Mann Whitney U Test, Sign Test, and Kruskal-Wallis H Test were conducted to compare RT across groups and time points. Additionally, an exploratory Spearman correlation was conducted to explore the relationship between RT and accuracy, as measured by the Correct Detection ratio. RT was calculated on trials where participants successfully pressed the space bar on Correct Detection targets. Eight participants, while having engaged with the task, failed to successfully detect any Correct Detection trials, so there was no RT data to represent Correct Detection RT ($N = 163$). Additionally, for all RT analyses involving the full vigilance task, data from two additional participants was excluded as outliers (>3 *SD*s above or below the mean and consistent with the most extreme outlying scores based on the IQR) ($N = 161$).

*Experimental vs. Control.* A Mann Whitney U Test was conducted to examine RT differences between the control and the combined experimental groups. Results showed a significant RT difference between the control group (*Mdn* = 658.06, Mean Rank = 107.26) and the combined experimental group (*Mdn* = 593.20, Mean Rank = 77.49), $U = 850.00$, $z = -2.62$, $p = .009$. This finding indicates that participants in the combined experimental group, who received some form of signal word/text highlight color motivational message, had significantly quicker RTs compared to participants in the control group.

*First Half vs. Second Half.* A Sign Test was conducted to compare RT between the first half of the vigilance task and the second half of the vigilance task. A Sign Test was used for this analysis because the data's distribution violated the Wilcoxon's Sign-Rank Test's assumption that the distribution of the differences between the two groups are symmetrical. Results showed no significant RT differences between the first half (*Mdn* = 602.76) and the second half of the task (*Mdn* = 590.44), after the manipulation was introduced, with 74 negative differences, 78 positive differences, and 0 ties, $z = -0.24$, $p = .808$.

*Signal Word/Text Highlight Color Comparison.* A Kruskal-Wallis H Test was run to assess RT differences across the signal word/text highlight color groups. The median and mean rank scores for each group were as follows: Group 1 (control, *Mdn* = 658.06, Mean Rank = 107.26), Group 2 (orange/WARNING, *Mdn* = 574.65, Mean Rank = 71.76), Group 3 (yellow/WARNING, *Mdn* = 591.57, Mean Rank = 74.84), Group 4 (no color/WARNING, *Mdn* = 615.94, Mean Rank = 91.46), Group 5 (orange/CAUTION, *Mdn* = 582.95, Mean Rank = 77.24), Group 6 (yellow/CAUTION, *Mdn* = 574.95, Mean Rank = 64.83), Group 7 (no color/CAUTION, *Mdn* = 624.50, Mean Rank = 86.43). Results indicated no significant differences between any of the groups, $\chi^2(6) = 11.94$, $p = .063$.

*Correlation - RT and Accuracy.* An exploratory Spearman correlation was conducted to assess the association between RT and vigilance task accuracy, as measured by the Correct Detection ratio. Results indicated a significant negative relationship between the two variables, $r(159) = -.51$, $p < .001$. This finding suggests that quicker RTs on the vigilance task are associated with a greater number of correct detections.

***Boredom***. Gender and ethnicity differences for boredom were evaluated using one-way ANOVAs with Bonferroni post hoc comparisons. No significant gender or ethnicity differences were found for either variable. An independent samples t-test was conducted to compare state boredom differences between participants who received the motivational message/signal word/alert symbol and participants who received no message. Additionally, a one-way ANCOVA was conducted to analyze the combined effects of varying signal words and text highlight colors on state boredom, while controlling for the effects of trait boredom. Further, Pearson and Spearman correlations were used to measure trait/state boredom's relationships with each other, and with RT/vigilance task accuracy (Correct Response and Correct Detection ratios).

*Full Vigilance Task - Boredom.*

*Experimental vs. Control.* An independent samples t-test was conducted to examine state boredom differences between the control and the combined experimental groups (H5 – Exp. 1). Results showed no significant differences between the control group ($M$ = 126.7, $SD$ = 23.7) and the experimental group ($M$ = 118.7, $SD$ = 30.3), $t(169) = 1.20$, $p = .231$, contrary to Hypothesis 5 – Exp. 1, which stated that "Participants who receive the motivational message in conjunction with a signal word and alert symbol will have significantly lower state boredom compared to participants who do not receive a motivational message."

*Signal Word/Text Highlight Color Comparison.* A one-way ANCOVA was conducted to analyze the combined effects of varying signal word and text highlight color on state boredom, while controlling for the effects of trait boredom (H6 – Exp. 1). First, the homogeneity of regression slopes assumption was evaluated by exploring the interaction between boredom proneness and signal word/text highlight color group. Results of the test indicated that the interaction was not significant, $F(6, 157) = 0.37$, $p = .900$. Therefore, the homogeneity of regression slopes assumption was met. Additionally, Levene's test for homogeneity of variances was met, $F(6, 164) = 0.78$, $p = .586$. No significant state boredom differences were seen between the signal word/text highlight color groups when controlling for trait boredom, $F(6,163) = 1.45$, $p = .199$, partial $\eta^2 = .05$, contrary to the hypothesis. The adjusted mean for each group was: Group 1 (control, $M$ = 128.7), Group 2 (orange/WARNING, $M$ = 115.2), Group 3 (yellow/warning, $M$ = 113.9), Group 4 (no color/WARNING, $M$ = 121.0), Group 5 (orange/CAUTION, *M* = 112.8), Group 6 (yellow/CAUTION, M = 126.9), Group 7 (no color/CAUTION, *M* = 121.3). The unadjusted mean scores were: Group 1 (control, *M* = 126.7, *SD* = 23.7), Group 2 (orange/WARNING, *M* = 113.9, *SD* = 33.0), Group 3 (yellow/WARNING, *M* = 115.1, *SD* = 30.1), Group 4 (no color/WARNING, *M* = 123.5, *SD* = 26.8), Group 5 (orange/CAUTION, *M* = 115.8, *SD* = 31.3), Group 6 (yellow/CAUTION, *M* = 123.4, *SD* = 27.7), Group 7 (no color/CAUTION, *M* = 121.7, *SD* = 34.5). This finding indicates that after controlling for the impact of trait boredom, varying the combined factor of signal word/text highlight color had no significant effect on state boredom. However, the covariate, trait boredom, was significantly associated with state boredom, $F(1,163) = 56.80$, $p < .001$, partial $\eta^2 = .26$, suggesting that trait boredom is a predictor of state boredom. Overall, this finding did not support Hypothesis 6 – Exp. 1, which stated that "Participants who receive any of the above-described signal word/text highlight color message condition combinations will have lower self-reported state boredom scores compared to participants in the control condition."

*Correlations - Trait Boredom, State Boredom, Accuracy, and RT.* A Pairwise Mann Whitney U Test and Kruskal-Wallis H Test were conducted to test for gender and ethnicity differences in trait and state boredom. No significant gender or ethnicity differences were found for either boredom variable. Additionally, Pearson and Spearman correlations were conducted to assess boredom's association with vigilance task accuracy and RT. A Pearson correlation was conducted to assess the relationship between trait and state boredom. Results showed a significant positive association between trait and state boredom, $r(169) = .49$, $p < .001$. This finding suggests that individuals who had high trait boredom levels also tended to report high state boredom levels. Additionally, exploratory Spearman correlations were conducted to examine trait and state boredom's association with RT. Results showed no significant relationship between trait boredom and RT, $r(159) = .15$, $p = .061$, or between state boredom and RT, $r(159) = .12$, $p = .127$. Additionally, Spearman correlations were conducted to examine the associations between trait boredom, state boredom, and vigilance task accuracy (as measured by the Correct Response ratio and Correct Detection ratio) (H7 – Exp. 1, H8 – Exp. 1). Results showed a significant negative relationship between trait boredom and the Correct Response ratio, $r(169) = -.22$, $p = .005$. Similarly, a significant negative correlation emerged between trait boredom and the Correct Detection ratio, $r(169) = -.21$, $p = .005$. These findings suggest that as trait boredom scores increased, vigilance task accuracy decreased. Further, a negative association was found between state boredom and the Correct Response ratio, $r(169) = -.26$, $p < .001$. A similar significant result was found for the association between state boredom and the Correct Detection ratio, $r(169) = -.25$, $p < .001$. These findings indicate that as state boredom scores increased, vigilance task accuracy decreased. These findings supported Hypotheses 7 and 8 – Exp. 1, which stated that "Participants with higher trait boredom scores will have lower total score accuracy on the sustained attention task" and "Participants with higher state boredom scores will have lower total score accuracy on the sustained attention task." Figures 6-9 visually depict the negative relationships between trait boredom, state boredom, and vigilance task accuracy.

**Figure 6**

*Scatterplot Depicting Relationship Between Trait Boredom and Correct Response Ratio. Y-axis begins at 0.90 to highlight visual differences.*

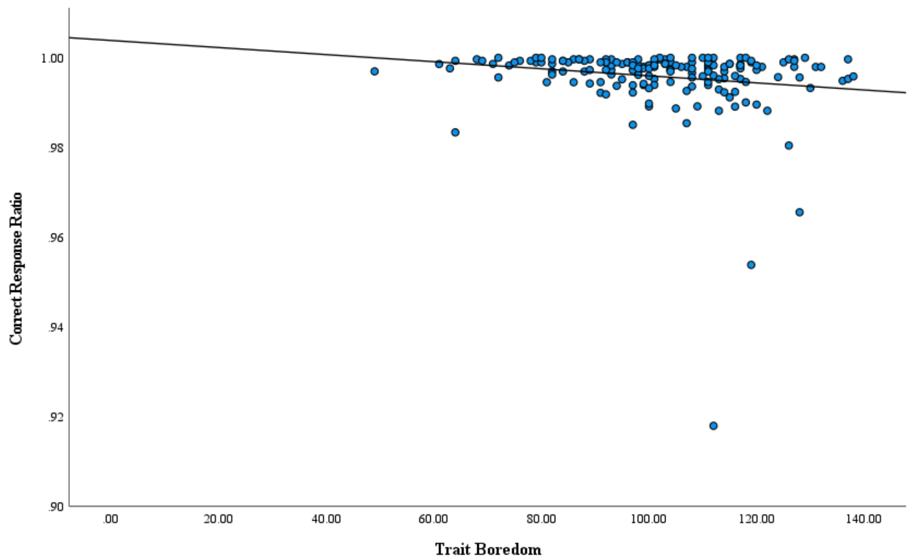



**Figure 7**

*Scatterplot Depicting Relationship Between State Boredom and Correct Response Ratio. Y-axis begins at 0.90 to highlight visual differences.*

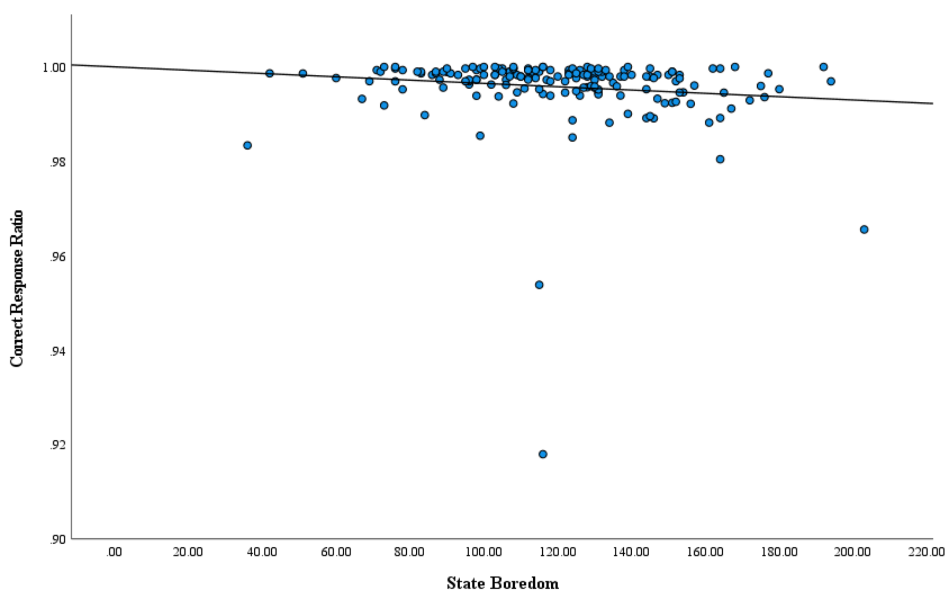


**Figure 8**

*Scatterplot Depicting Relationship Between Trait Boredom and Correct Detection Ratio.*

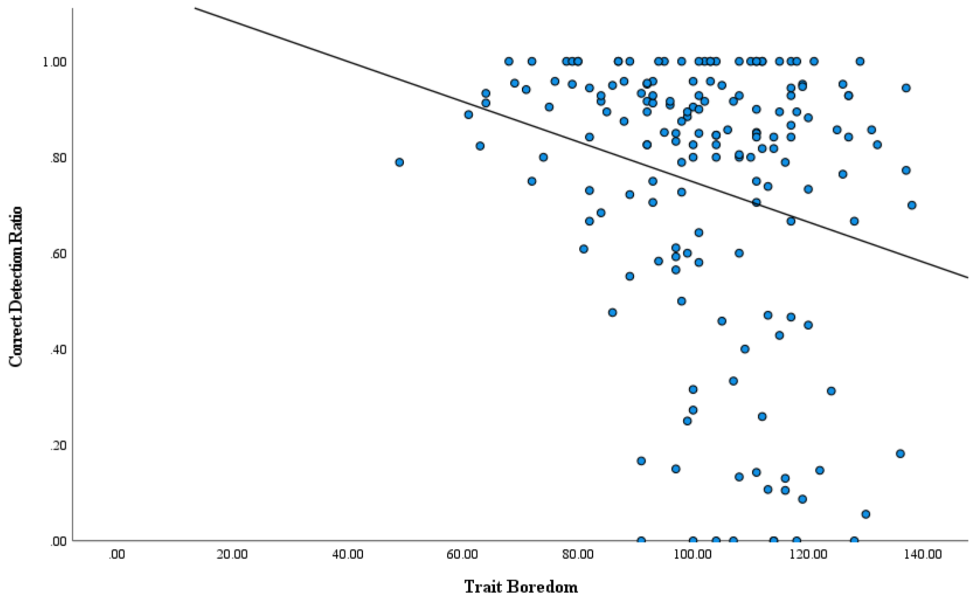
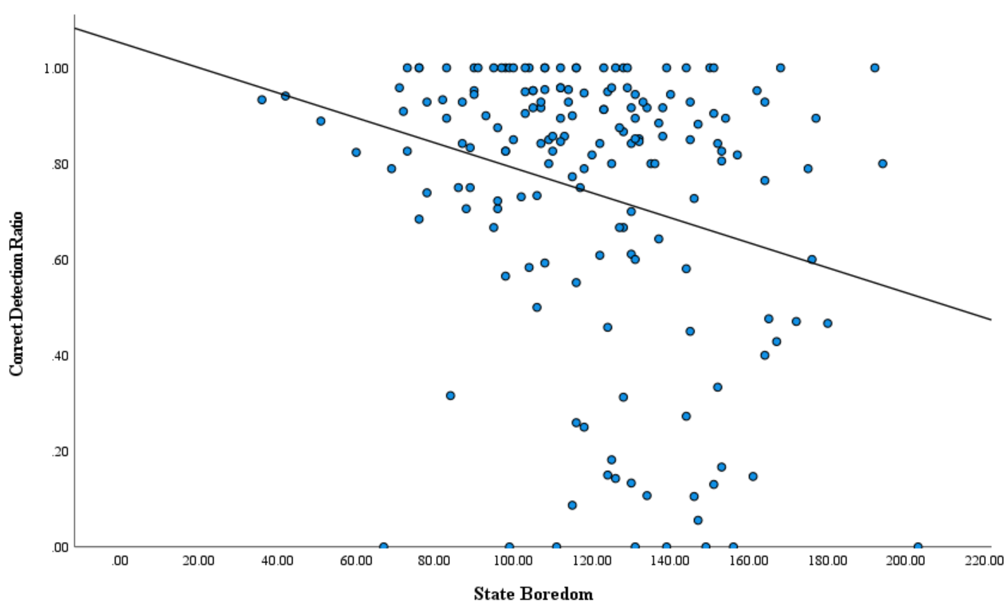



**Figure 9**

*Scatterplot Depicting Relationship Between State Boredom and Correct Detection Ratio.*



## Discussion

Experiment One sought to investigate the effectiveness of using varying signal words paired with incentivizing messages to improve performance on vigilance-based tasks, as measured by overall performance accuracy and RT. Overall, the findings suggest that participants who received the motivating message (e.g., message that denotes hazard) in conjunction with a signal word and alert symbol had higher accuracy on the vigilance-based task compared to participants who received no message, but varying the signal word did not change the impact of the message in terms of performance outcomes. Additionally, the results demonstrated associations between boredom, RT, and accuracy on the vigilance task, where lower boredom levels/faster RT were related to better accuracy on the vigilance task. Table 4 lists all Experiment One hypotheses and indicates whether support was found in the current study. Further, the following paragraphs discuss the key findings of Experiment One.

**Table 4**

*Hypotheses - Experiment One*

| Hypothesis # | Hypothesis | Support (Yes/No) |
|---|---|---|
| H1 | Performance accuracy on the sustained attention task will improve within subjects when presented with a manipulation (signal word/text highlight color combination) compared to performance on a training block where no manipulation is present. | Yes |
| H2 | Participants who receive the motivational message in conjunction with a signal word and alert symbol will perform significantly more accurately on the task compared to participants who do not receive a motivational message. | Yes |
| H3 | Performance accuracy on the sustained attention task will decrease within subjects from the first half of the task compared to the second half of the task. | No |
| H4 | Significant differences will be seen between the combined signal word/text highlight color conditions. Specifically, participants who receive the combined orange/WARNING prompt will have the highest accuracy on the sustained attention task, followed by participants who receive the combined yellow/CAUTION prompt, and those in the control condition will have the lowest accuracy out of any of the conditions. | Partial Support |
| H5 | Participants who receive the motivational message in conjunction with a signal word and alert symbol will have significantly lower state boredom compared to participants who do not receive a motivational message. | No |
| H6 | Participants who receive any of the above-described signal word/text highlight color message combinations will have lower self-reported state boredom scores compared to participants in the control condition. | No |
| H7 | Participants with higher trait boredom scores will have lower total score accuracy on the sustained attention task. | Yes |
| H8 | Participants with higher state boredom scores will have lower total score accuracy on the sustained attention task. | Yes |

In the current study, when presented with different signal word/text highlight color variations (no color/no signal word, orange/WARNING, yellow/WARNING, no color/WARNING, orange/CAUTION yellow/CAUTION, no color/CAUTION) in conjunction with the motivational message denoting the consequence of noncompliance ("You must complete the following task with at least **95%** accuracy, or you will be required to re-start the activity.") and an alert symbol, significant accuracy differences, as measured by the Correct Detection ratio, were seen across several of the signal word/text highlight color groups. The Correct Detection ratio measured performance accuracy by calculating the proportion of correct detections (instances where the clock hand jumped, and the participant pressed the space bar) out of total jump (target) trials. In this study, the control group, who received no signal word/text highlight color combination or motivational message, performed significantly less accurately than participants who received the orange WARNING message, the yellow WARNING message, the WARNING message with no text highlight color, and the yellow CAUTION message. However, no significant accuracy differences were observed between the control and the following experimental groups: orange CAUTION, CAUTION message with no text highlight color. Additionally, there were no significant Correct Detection ratio differences seen between the signal word/text highlight color groups themselves. These findings provide partial support for the current study's hypotheses, suggesting that the presentation of a motivating message in conjunction with a signal word and alert symbol prior to beginning a vigilance task may increase the number of correct detections to a target compared to performance when no message is shown before the start of the task. However, it is unclear why significant accuracy differences were not seen between the control group and the other signal word/text highlight color groups. Additionally, because there were no significant Correct Detection ratio differences between any of the signal word/text highlight color groups themselves, it appears that varying the signal word did not impact performance. Rather, the presence of the motivating message in conjunction with a signal word and alert symbol was effective in improving performance outcomes as measured by the number of correct detections compared to when participants were shown no message at all.

Moreover, when performance accuracy was measured by the Correct Response ratio, no significant differences were seen between any of the groups. The Correct Response ratio was a proportion of total number of "correct" trials, with "correct" meaning that the participant responded appropriately to that given trial (ex. correctly pressed the space bar on jump trials and did not press the spacebar on non-jump trials). Further, no state boredom or RT differences were observed between any of the groups. Taken together, these findings suggest that in the context of a basic vigilance task, varying signal words/text highlight colors is not effective in capturing attention or improving performance. However, participants who viewed the motivational message combined with the signal word and alert symbol had higher accuracy, as measured by the Correct Detection ratio, compared to participants in the control group who received no message prior to the task. More research is needed to understand the individual influence of the motivational message with and without corresponding signal words and alert symbols, as the current study did not allow for the measurement of each component's individual influence (motivational message, signal word, alert symbol). However, it is clear that varying the signal word and text highlight color did not alter performance outcomes in participants who viewed any arrangement of the message. While this finding was contrary to our hypotheses, other research supports the lack of findings in the current study. Specifically, opinions vary on whether WARNING and CAUTION and their respective text highlight colors (orange and yellow) cause users to reliably distinguish between hazards and consider their behaviors accordingly. Previous research has provided evidence that in workplace environments, users do not reliably differentiate between WARNING and CAUTION in safety hazards (Mayhorn et al., 2015). Additionally, other research has shown that most people do not reliably distinguish between hazards depicted using orange and yellow (Chapanis, 1994; Mayhorn et al., 2004; Wogalter et al., 1998a). The findings of the current study may support this research, suggesting that users do not reliably differentiate between WARNING and CAUTION or orange and yellow on sustained attention tasks. However, more research is needed to test this hypothesis.

As described in previous research, motivation is highly influential in shaping behavioral outcomes (Wogalter, 2006b), and extrinsic motivation/incentivization is associated with less of a vigilance decrement (Garner et al., 2024). Therefore, to further test the effect of the motivational message itself ("If you do not complete the following task with at least **95%** accuracy, you will be required to re-start the activity.") when presented alongside signal words and alert symbols, accuracy of participants in the control group was compared to accuracy of participants in the combined experimental groups, as measured by both the Correct Response ratio and Correct Detection ratio. Findings showed that participants who received the motivational message in conjunction with a signal word and alert symbol performed more accurately, as measured by both accuracy ratios, and had faster RTs on the vigilance task than participants who received no message prior to completing the task. Previous research shows that incentivization increases accuracy on sustained attention tasks (Garner et al., 2024; McGough & Mayhorn, 2022), and RT is a metric often considered to be an indicator of attention (Yamashita et al., 2021). Therefore, this finding provides additional support that utilizing motivational messages in conjunction with signal words and alert symbols may be an effective strategy to increase accuracy on vigilance-based tasks. However, no state boredom differences were observed between the combined experimental group and the control group, contrary to hypotheses. However, it is possible that in the current study, the message manipulation was not mitigating enough to reduce self-reported state boredom on the vigilance task, which was designed to be monotonous and thus can be perceived as boring (Manly et al., 1999; Robertson et al., 1997).

While no significant accuracy differences were seen *between* signal word/text highlight color groups, our findings did provide evidence of within-subject accuracy differences before the signal word message was presented compared to after it was presented. To make this comparison, performance accuracy was compared in all experimental condition groups (not control group) between the ten-minute training block and the first ten minutes of the full 50-minute task, after the signal word message was presented. Results showed that participants performed more accurately, as measured by the Correct Response ratio, on the vigilance task once they received the signal word message compared to in the training block, providing evidence that the signal word message presentation was initially effective in re-engaging and/or motivating participants to the task. The goal of signal words and warning systems are to capture a user's attention/alert them of the potential for an undesirable outcome (ANSI, 2002; Mayhorn et al., 2015) and providing a warning signal intended to warn of a consequence is a type of extrinsic motivator (Ryan & Deci, 1985). This finding provides support for the effectiveness of the combined motivational message, signal word, and alert symbol, such that the message was impactful in capturing participants' attention and improving accuracy within the first ten minutes of its presentation. However, it is also possible that observed changes were influenced by or due to practice effects, where the participant became familiar with the task over time. To explore this theory further, an additional exploratory analysis was conducted to test for performance differences in the control group, who did not receive the message manipulation. Results in this supplementary analysis indicated a significant difference, where participants had increased accuracy in the first ten minutes compared to the control. Therefore, it is likely that practice effects influenced the observed differences in the training analysis to some extent.

Contrary to hypotheses, no vigilance decrement was observed when comparing the accuracy (as measured by the Correct Response ratio and Correct Detection ratio) during the first half of the 50-minute task to the second half of the task. Additionally, no vigilance decrement was observed between the first and second halves of the tasks when performance was measured by RT, a metric that is often considered an indicator of attention on vigilance tasks (Yamashita et al., 2021). While the overarching aim of this study was to explore the potential for signal words and incentives to increase accuracy on vigilance tasks, it was still anticipated that a vigilance decrement would be seen over time due to the cognitive demands of the task (Hancock & Warm, 1989; Hockey, 1997; Parasuraman et al., 1987). One possible explanation for these results is that the presence of the signal word motivational messages mitigated the vigilance decrement to some extent, resulting in less of an accuracy decrease over time. This finding would support other research that shows that providing incentives for performance can reduce evidence of the vigilance decrement on sustained attention tasks (Garner et al., 2024). However, when performing additional exploratory analyses to assess this idea, a vigilance decrement was still not observed when comparing performance on the first half and second half in only the control group. Therefore, it is unlikely that the presence of the signal word message mitigated signs of the vigilance decrement between the two halves. Other possible explanations for the lack of observed vigilance decrement could be the testing environment, where many participants likely completed the study in a quiet, comfortable space in their homes. This factor may have minimized distractions and allowed participants to remain vigilant for the full duration of the task. Additionally, all participants completed the study for SONA credit as part of a course requirement and this may have influenced behavior. Specifically, completing the study to earn SONA credits may have been extrinsically motivating for some participants, perhaps affecting performance and resulting in a lack of vigilance decrement between the two halves of the task (Garner et al., 2024; Robison & Nguyen, 2023; Deci & Ryan, 1985; Ryan & Deci, 2000). Finally, it is possible that the vigilance decrement could have been detected if the comparison was made across different intervals. For example, instead of comparing accuracy between the first half and second half of the task, perhaps a vigilance decrement would be more apparent if measured across additional, more lengthy time intervals.

An additional component of Experiment One explored the relationship between trait boredom, state boredom, and vigilance-task accuracy. Results provided evidence in support of the hypotheses, where higher trait and state boredom levels were associated with lower accuracy on the vigilance task, as measured by both the Correct Response ratio and Correct Detection ratio. This finding supports Mindlessness Theory that posits vigilance decrements are due to the monotony of tasks that leads to mind-wandering and experiences of boredom, eventually contributing to an increase in errors (Manly et al., 1999; Robertson et al., 1997). Additionally, these findings support previous research that demonstrates reduced sustained attention task accuracy is associated with high levels of both trait (Hunter & Eastwood, 2018; Malkovsky et al., 2012) and state boredom (Hunter & Eastwood, 2018; Petranker & Eastwood, 2021). As such, the impact of boredom (both in-the-moment and predisposition toward) is important to consider when evaluating performance on sustained attention tasks, as high experiences of boredom will likely result in increased errors on tasks that require sustained attention. RT’s association with correct detection performance accuracy and boredom was also examined in an exploratory analysis. Results indicated no relationship between trait boredom, state boredom, and RT. However, a significant association was found between RT and accuracy, such that faster RTs were associated with more correct detections. This finding supports research conducted by Morais et al. (2024), providing additional evidence that faster RTs are correlated with a greater number of correct detections.

# Experiment Two

Experiment Two sought to examine the effects of social motivation (gamification and reward incentivization) on accuracy and RT on a basic vigilance task. Additionally, like Experiment One, this study explored the relationships between trait boredom, state boredom, RT, and accuracy on a basic vigilance task. The central aim of Experiment Two was to complement the findings of Experiment One, to further inform the design of user-centered protocols requiring high sustained attention. Specifically, Experiment Two focused on understanding the effectiveness of social motivation via various competition formats as a tool to improve performance on vigilance tasks, as these strategies may have potential to improve efficiency and reduce errors across a range of task types.

## Method

### *Participants*

Data collection for Experiment Two took place remotely from August 2024 through December 2024. All participants were undergraduate students enrolled in an Introduction to Psychology course at North Carolina State University. Participants were recruited through this course, as a syllabus requirement was to select between participating in research or completing an alternative written assignment of equal effort. All participants were required to have access to a computer with a keyboard and internet connection to complete the study. This study's procedures were approved by an institutional review board (IRB) before data collection began. Prior to beginning the study, participants read through and acknowledged informed consent documentation.

A power analysis was conducted using G* Power 3 software (Faul et al., 2007) to determine the necessary sample size. The analysis was conducted based on the Kruskal-Wallis H test, the statistical test that required the largest sample size among the analysis plan, to ensure adequate power for all tests. However, the software does not provide a built-in option for the Kruskal-Wallis H test, so the analysis was conducted using a one-way ANOVA model as an approximation. With an effect size of .30 and power of .80, at least 128 participants were necessary for analyses to be adequately powered. Ultimately, 170 participants completed Experiment Two. However, seven participants were excluded from analyses due to data that suggested a lack of engagement with the task (never pressed the space bar during the 50-minute full duration of the task) ($n$ =2), a low number of task trials (<1700 total trials) ($n$ = 3), or for outlying scores on the Mackworth Clock Test Total Accuracy (Correct Response Ratio) (>3 *SD*s below the mean after removing all other data excluded from analyses and consistent with the most extreme outliers based on the Interquartile Range (IQR)) ($n$ = 2). Note that only the two most extreme outliers were excluded from analyses because they were indicative that the participant did not appropriately complete the task. However, less extreme outliers (datapoints that were within three *SD*s of the mean but were identified as outliers based on the IQR) remained in the analyses, as they appeared to be true outliers based on performance and not based on data noise. Thus, the final sample consisted of 163 participants (Seventy-seven participants identified as male, 83 participants identified as female, 2 participants identified as other genders not listed, and 1 participant selected that they preferred not to answer. Ages ranged from 18 to 26 ($M$ = 18.99 $SD$ = 1.33), and participants identified as the following ethnicities: Black/African American (4.3%), Asian American/Pacific Islander (9.8%), White/Caucasian (73.0%), Hispanic/Latino (4.9%), American Indian/Alaskan Native (1.2%), multiple ethnicities (4.3%), Prefer not to answer (2.5%).

***Design***

The current study used a 4x2 mixed factorial design to explore the effects of social motivation on vigilance task accuracy, RT, and boredom. Social motivation, the first independent variable, was manipulated between-subjects across four groups: a control group, gamification ranked leaderboard, reward incentivization, and a combined group (Group 1 - control: $n$ = 38, Group 2 - leaderboard: $n$ = 34, Group 3 – reward incentivization: $n$ = 41, Group 4 - combined: $n$ = 50). The second independent variable, time of measurement, was manipulated within-subjects before the manipulation was introduced (directly after the ten-minute training block) compared to after the manipulation was introduced (ten minutes after the manipulation was presented). Consistent with Experiment One, task accuracy, RT, and state boredom were measured as dependent variables with comparisons made between the first and second half of the vigilance task. Additionally, trait boredom was measured as a covariate.

***Independent Variables***

*Social Motivation.*

*Control.* In the control condition for Experiment Two, participants did not receive any social motivation manipulation. Rather, participants began immediately with the Mackworth Clock Test after the training without viewing the ranked leaderboard or reward incentivization competition formats.

*Ranked Leaderboard.* A researcher-designed ranked leaderboard was shown to participants in the gamification condition for Experiment Two. This leaderboard displayed a ranking of other participants' scores on the upcoming vigilance task based on their performances on the same task. This leaderboard was a manipulated variable intended to motivate accurate performance on the task. Very weak deception, as approved by the IRB, was used, as the leaderboard showing participant high scores was not representative of any real participant data. Figure 10 shows an image of what participants viewed in the gamification ranked leaderboard and combined conditions.

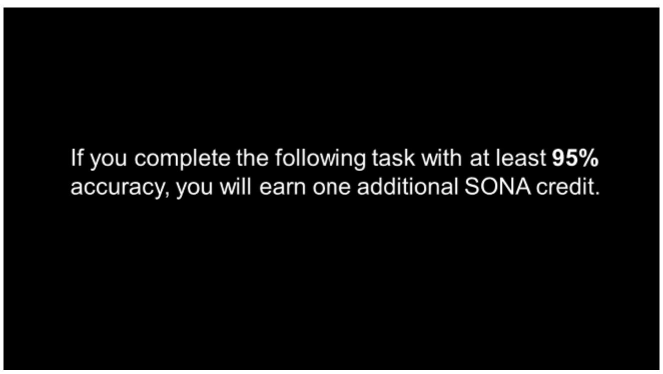
**Figure 10**

*Ranked Leaderboard*


*Reward Incentivization.* In the reward incentivization condition for Experiment Two, participants viewed a prompt that read "If you complete the following task with at least **95%** accuracy, you will receive one additional SONA credit." This message was a manipulated variable that was intended to motivate performance accuracy on the vigilance task. Each semester, participants enrolled in an Introduction to Psychology course are required to either participate in research or complete an alternative assignment of equal effort. For participants who elect to partake in research, they are required to earn a pre-determined number of SONA credits. Each study listed on the SONA platform is worth a certain number of SONA credits, depending upon the length of the study. So, the goal of the reward incentivization message was to serve as a motivator for participants to perform accurately on the task to earn an additional SONA credit that could count towards their total credits needed for the course requirement. Very weak deception was used with this manipulation, as approved by the IRB, because all participants who completed the study in full received an additional SONA credit, regardless of their performance accuracy. To maintain equal SONA credits for each study condition, all participants who completed the study in full received one additional SONA credit, regardless of whether they received this message or not. Figure 11 shows an image of what participants viewed in the reward incentivization and combined conditions.

**Figure 11**

*Reward Incentivization Message*




*Combined.* In the Combined condition for Experiment Two, participants received both manipulations. First, participants viewed the reward incentivization message that read "If you complete the following task with at least **95%** accuracy, you will receive one additional SONA credit." On the following screen, participants then viewed the ranked leaderboard that displayed a ranking of other participants' scores on the upcoming vigilance task based on their performances on the same task.

### *Measures*

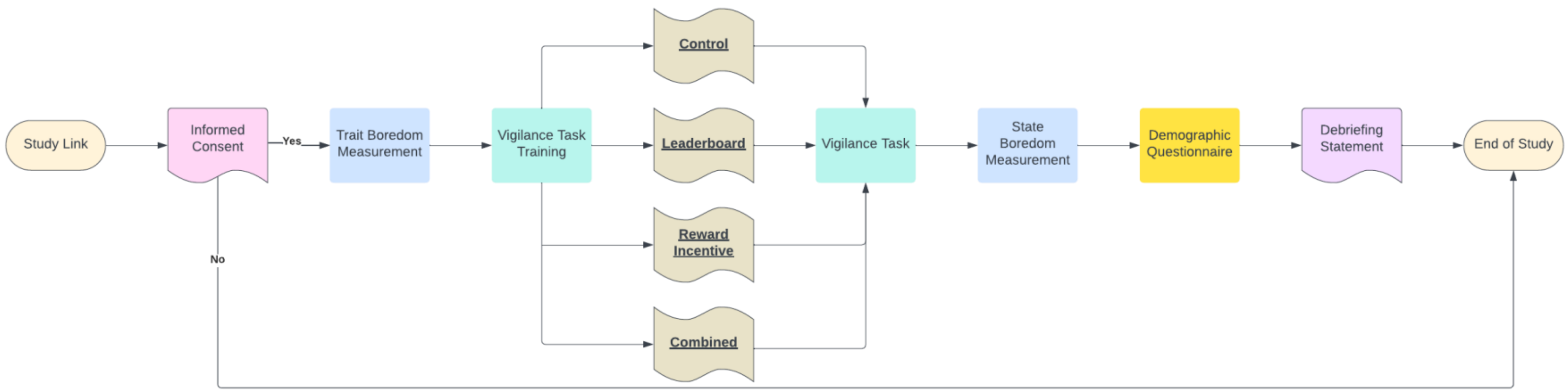
All dependent measures were identical to Experiment One, such that the Mackworth Clock Test (Mackworth, 1948) was used to measure accuracy (correct responses, correct detections) and RT. Additionally, like in Experiment One, the Boredom Proneness Scale (BPS; Farmer & Sundberg, 1986) and Multidimensional State Boredom Scale (MSBS; Fahlman et al., 2013) were used to collect information about trait and state boredom. For the current sample, the Cronbach's alpha indicated good reliability for the BPS ($\alpha = .835$) and excellent reliability for the MSBS ($\alpha = .963$). Lastly, the same Demographic Questionnaire was administered as in Experiment One. The only measurement that was not collected in Experiment Two was the modified Ishihara Test (Ishihara, 1917), as it was not relevant to the research questions of Experiment Two.

### *Procedure*

All participants received a link for PsyToolkit upon signing up for the study. PsyToolkit is an online repository for psychological experiments and data collection (Stoet 2010, 2017). Participants first read through and acknowledged the informed consent documentation. Following the informed consent, participants were randomly assigned to one of four social motivation conditions. In a procedure similar to Experiment One, participants then completed the same order of tasks: Boredom Proneness Scale (BPS; Farmer & Sundberg, 1986), Mackworth Clock Test Training, Mackworth Clock Test (Mackworth, 1948), Multidimensional State Boredom Scale (MSBS; Fahlman et al., 2013), and Demographic Survey. After the study, all participants viewed a debriefing statement with an option available to download the form for their records. In total, the study took approximately 80-90 minutes to complete. Figure 12 shows a visual depiction of the procedural flow for Experiment Two.

**Figure 12**

*Procedural Flow for Experiment Two*



### ***Statistical Analyses***

All data was evaluated for normality by examining the skewness and kurtosis levels (>2), visually inspecting histograms of the distributions, and examining the Shapiro Wilks and Kolmogorov-Smirnov Tests. The trait boredom (BPS) and state boredom (MSBS) variables were normally distributed. However, the accuracy variables for the Mackworth Clock Test training, Correct Response ratio, and Correct Detection ratio were negatively skewed. Additionally, the RT variables had non-normal distributions. Therefore, nonparametric analyses were performed for all analyses involving the vigilance task accuracy and RT variables, and any analyses that focused solely on boredom variables were parametric. All data was handled consistently with the manner it was analyzed in Experiment One, except the signal word/text highlight color manipulation was replaced with the social motivation manipulation. Table 5 and 6 outline the descriptive statistics for the sample.

**Table 5**

*Descriptive Statistics - Mackworth Clock Test Accuracy, Trait Boredom, State Boredom*

| | Total | | | | Control | | | | Experimental | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *N* | *M* | *SD* | *Mdn* | *N* | *M* | *SD* | *Mdn* | *N* | *M* | *SD* | *Mdn* |
| Training – Correct Responses | 163 | 0.968 | 0.102 | 0.993 | 38 | 0.976 | 0.096 | 0.995 | 125 | 0.966 | 0.104 | 0.992 |
| Correct Responses | 163 | 0.995 | 0.006 | 0.996 | 38 | 0.995 | 0.005 | 0.996 | 125 | 0.995 | 0.006 | 0.996 |
| Correct Detections | 163 | 0.579 | 0.322 | 0.667 | 38 | 0.573 | 0.332 | 0.683 | 125 | 0.581 | 0.320 | 0.667 |
| Incorrect Detections | 163 | 0.003 | 0.005 | 0.001 | 38 | 0.003 | 0.004 | 0.002 | 125 | 0.003 | 0.005 | 0.001 |
| Missed Detections | 163 | 0.420 | 0.322 | 0.333 | 38 | 0.426 | 0.332 | 0.317 | 125 | 0.419 | 0.320 | 0.333 |
| Trait Boredom | 163 | 100.939 | 18.280 | 102.000 | 38 | 101.184 | 17.596 | 102.000 | 125 | 100.864 | 18.552 | 102.000 |
| State Boredom | 163 | 115.847 | 36.025 | 118.000 | 38 | 116.526 | 36.410 | 118.500 | 125 | 115.640 | 36.053 | 118.000 |

*Note*. Correct Response ratios were calculated by subtracting missed detections and incorrect detections from the total number of trials and dividing that score by the total number of trials. Correct Detection and Missed Detection ratios were calculated out of the total number of jump (target) trials. The Incorrect Detection ratio was calculated out of the total number of trials.

**Table 6**

*Descriptive Statistics – Correct Detection RT*

| | *N* | *M* | *SD* | *Mdn* |
|---|---|---|---|---|
| Total | 151 | 647.53 | 94.62 | 634.95 |
| Control | 34 | 646.46 | 95.82 | 639.07 |
| Experimental | 117 | 647.84 | 94.68 | 629.46 |

*Note.* RT was calculated on trials where participants successfully pressed the space bar on Correct Detection targets. Further, twelve participants, while having engaged with the task, failed to successfully detect any Correct Detection trials, so there was no RT data to represent for Correct Detection RT.

**Results**

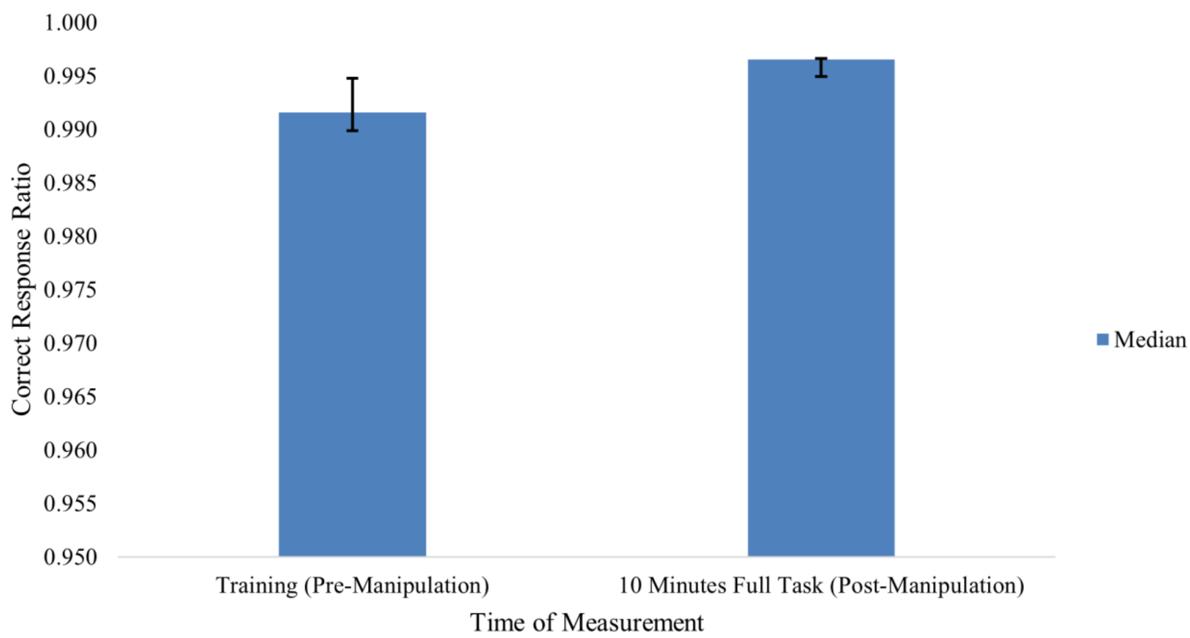
***Mackworth Vigilance Task Performance***. Kruskal-Wallis H Tests were conducted to evaluate gender and ethnicity differences based on vigilance performance. Results showed significant gender differences in RT on the Correct Detection ratios, $\chi^2(2) = 9.05$, $p = .011$. Of note, one participant selected the "preferred not to answer" option for the Gender Demographic question, but this option was not considered a separate gender category. Therefore, analyses were only performed comparing differences between the other gender options. Post hoc Dunn's Test with Bonferroni Corrections showed significant gender differences between participants who identified as male (*Mdn* = 622.50) and participants who identified as female (*Mdn* = 658.84), adj. $p = .009$, where participants who identified as male demonstrated significantly quicker RTs on correct detections than participants who identified as female. No significant gender differences were found in any of the other variables (Correct Detection ratio, Correct Response ratio, Training Accuracy ratio).

Additionally, four participants selected the "preferred not to answer" option for the Ethnicity Demographic question, but this option was not considered a separate ethnicity category. Therefore, analyses were only examined for differences between the other ethnicity options. No significant ethnicity differences were found in any of the variables (Correct Detection RT, Correct Detection ratio, Correct Response ratio, Training Accuracy ratio).

*Training.* A Sign Test was conducted to compare the Correct Response ratio on the vigilance task before and after the manipulation was introduced, as the data's distribution violated the Wilcoxon's Sign-Rank Test's assumption that the distribution of the differences between the two groups are symmetrical (H1 – Exp. 2). This analysis was only conducted comparing the Correct Response ratio score and not the Correct Detection ratio, as the number of jump trials in each ten-minute window contained extremely limited data points (~4 jump trials). For this analysis, the Correct Response ratio in the ten-minute training block was compared to the Correct Response ratio in the first ten minutes of the actual task, after the social motivation manipulation was introduced. Accuracy during the first ten minutes of the actual task was calculated in analysis by using the same number of trials as the participant completed in the training block to compare performance. As this analysis sought to compare Correct Response ratio differences based on the manipulation, participants in the control condition were excluded. Results indicated a significant difference in the Correct Response ratio between the training (*Mdn* = 0.992) and first ten minutes of the task (*Mdn* = 0.997), with 25 negative differences, 81 positive differences, and 19 ties, $z = -5.34$, $p < .001$. This finding suggests that participants performed significantly more accurately (as measured by overall correct responses) on the vigilance task after receiving a social motivation manipulation compared to in the training block, providing support for Hypothesis 1 – Exp. 2 that stated, "Performance accuracy on the sustained attention task will improve within subjects when presented with a social motivation manipulation compared to performance on a training block where no manipulation is present." Figure 13 visually depicts the Correct Response ratio differences between the pre-manipulation vigilance task training and the first ten minutes of the task post-manipulation.

**Figure 13**

*Correct Response Ratio Differences Between Training and First 10 Minutes of Full Task. Y-axis begins at 0.950 to highlight visual differences.*




*Note.* Error bars indicate bootstrapped 95% confidence intervals (CI), calculated from 5,000 samples.

*Full Vigilance Task – Correct Responses.*

*Experimental vs. Control.* A Mann-Whitney U test was performed to compare the vigilance task accuracy of participants who received the social motivation manipulations compared to participants in the control condition (H2 – Exp. 2: Correct Responses). For this analysis, the group that received the social motivation manipulation included the combined three experimental groups from the study, who all received some form of motivating element (ranked leaderboard, reward incentive, combined). When accuracy was measured using the Correct Response ratio, results indicated no significant differences between the control group (*Mdn* = 0.996, Mean Rank = 82.71) and participants who received the social motivation manipulation (*Mdn* = 0.996, Mean Rank = 81.78), contrary to hypotheses, $U = 2348.00$, $z = -0.11$, $p = .916$.

This finding does not provide support for Hypothesis 2 – Exp. 2: Correct Responses, which stated that "Participants who receive a social motivation manipulation will perform significantly more accurately on the task compared to participants who are in the control condition."

*First Half vs. Second Half.* A Wilcoxon-Sign Rank Test was conducted to compare performance accuracy, as measured by the Correct Response ratio, on the vigilance task between the first half of the task and the second half of the task (H3- Exp. 2: Correct Responses). Results showed no significant Correct Response ratio score differences between the first (*Mdn* = 0.996) and second halves (*Mdn* = 0.996) of the task, with 74 negative ranks, 68 positive ranks, and 21 ties, $z = -0.35$, $p = .728$. This finding was contrary to Hypothesis 3 – Exp. 2: Correct Responses, which stated that "Performance accuracy on the sustained attention task will decrease within subjects from the first half of the task compared to the second half of the task."

*Social Motivation Comparison.* A Kruskal-Wallis H Test was performed to assess the effects of social motivation on the vigilance task accuracy variables (H4 – Exp. 2: Correct Responses). When task accuracy was measured by the Correct Response ratio, the analysis showed significant accuracy differences between the groups, $\chi^2(3) = 10.21$, $p = .017$. Post hoc comparisons were made using Dunn's Test with Bonferroni Corrections. Results showed that Group 2 (ranked leaderboard) performed significantly less accurately than Group 3 (reward incentive), but no significant differences were screen between any of the other groups. The mean rank and median score for each group were as follows: Group 1 (control, *Mdn* = 0.996, Mean Rank = 82.71), Group 2 (Leaderboard, *Mdn* = 0.995, Mean Rank = 67.40), Group 3 (Reward Incentive, *Mdn* = 0.997, Mean Rank = 100.38), Group 4 (Combined, *Mdn* = 0.996, Mean Rank = 76.32). Table 7 outlines the pairwise comparisons from the post-hoc test. Taken together, these findings provide partial support for Hypothesis 4 – Exp. 2: Correct Responses, which stated the following: “Significant differences will be seen across the social motivation conditions. Specifically, participants will have the highest accuracy on the sustained attention task in the combined leaderboard/reward incentivization condition, where the most motivating stimuli are present, followed by participants in the standalone gamification leaderboard and reward incentivization conditions, and participants in the control condition will have the lowest accuracy out of any condition.”

**Table 7**

*Dunn's Multiple Comparison Test - Correct Response Ratio*

| Comparison | Group | Test Statistic | Std. Error | Std. Test Statistic | *p* | Adj. *p*[a] |
|---|---|---|---|---|---|---|
| 2-4 | Leaderboard vs. Combined | -8.923 | 10.491 | -0.851 | .395 | 1.000 |
| 2-1 | Leaderboard vs. Control | 15.313 | 11.141 | 1.374 | .169 | 1.000 |
| 2-3 | Leaderboard vs. Reward Incentive | -32.981 | 10.947 | -3.013 | .003 | .016* |
| 4-1 | Combined vs. Control | 6.391 | 10.157 | 0.629 | .529 | 1.000 |
| 4-3 | Combined vs. Reward Incentive | 24.058 | 9.944 | 2.419 | .016 | .093 |
| 1-3 | Control vs. Reward Incentive | -17.668 | 10.628 | -1.662 | .096 | .579 |

a. Bonferroni correction, *$p < .05$

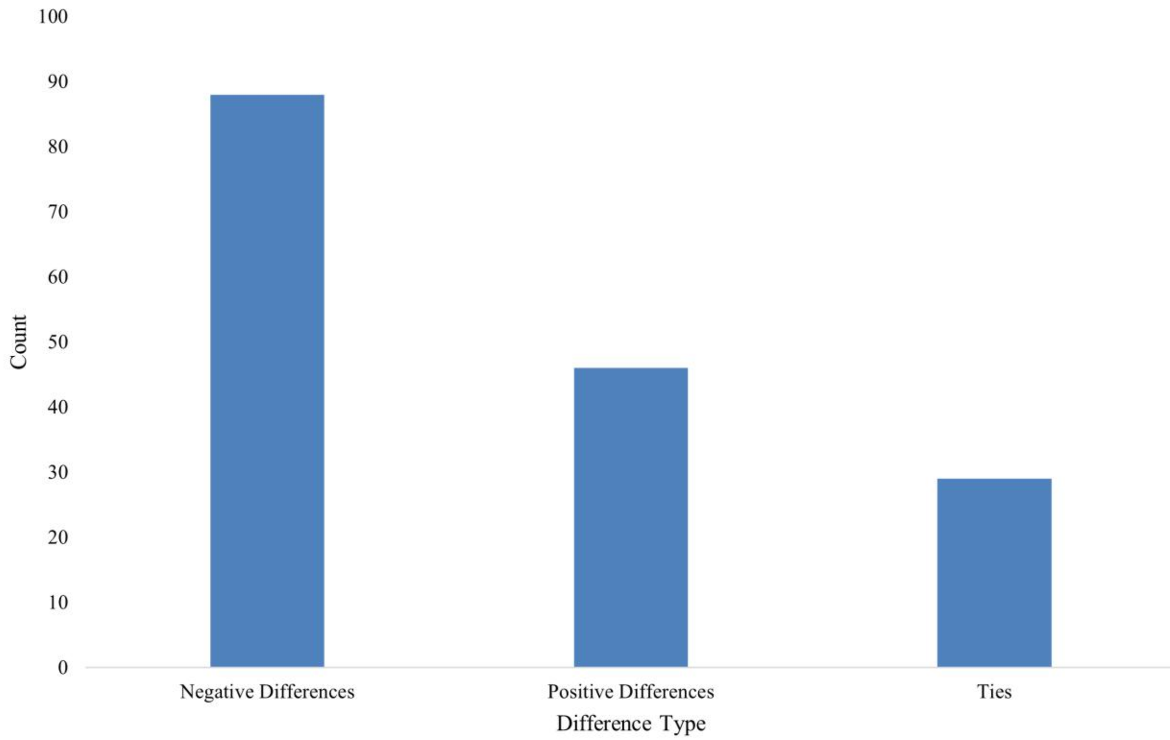
*Full Vigilance Task – Correct Detections.*

*Experimental vs. Control.* A Mann-Whitney U test was performed to compare the vigilance task accuracy of participants who received the social motivation manipulations compared to participants in the control condition (H2 – Exp. 2: Correct Detections). For this analysis, the group that received the social motivation manipulation was the combined three experimental groups from the study, who all received some form of motivating element (ranked leaderboard, reward incentive, combined).When accuracy was measured using the Correct Detection ratio, no significant differences were seen between the control (*Mdn* = 0.683, Mean Rank = 80.97) and combined experimental group (*Mdn* = 0.667, Mean Rank = 82.31), $U$ = 2336.00, $z = -0.15$, $p = .878$. This finding does not provide support for Hypothesis 2 – Exp. 2: Correct Detections, which stated that “Participants who receive a social motivation manipulation will perform significantly more accurately on the task compared to participants who are in the control condition."

*First Half vs. Second Half.* A Sign Test was conducted to compare performance accuracy between the two halves of the vigilance task when measured by the Correct Detection ratio (H3-Exp. 2: Correct Detections), as the data's distribution violated the Wilcoxon's Sign-Rank Test's assumption that the distribution of the differences between the two groups are symmetrical. Results indicated significant Correct Detection ratio score differences between the first and second half of the task. The median scores were nearly identical ($Mdn$ = 0.667), but the analysis revealed significant differences in the distribution of scores, $z = -3.54$, $p < .001$, with 88 negative differences, 46 positive differences, and 29 ties. This finding indicates that the Correct Detection ratio was significantly lower in the second half of the vigilance task compared to the first half, providing support for Hypothesis 3 – Exp. 2: Correct Detections that stated, "Performance accuracy on the sustained attention task will decrease within subjects from the first half of the task compared to the second half of the task." Therefore, support was garnered for Hypothesis 3 – Exp. 2 when vigilance task accuracy was measured by the Correct Detection ratio but not by the Correct Response ratio. Figure 14 visually depicts the Correct Detection ratio negative differences, positive differences, and ties between the first and second halves of the task.

**Figure 14**

*First vs. Second Half Correct Detection Ratio Accuracy: Sign Test Results*




*Note.* Negative Differences indicates the count of participants who had a lower Correct Detection ratio in the second half of the task than in the first. Positive Differences indicates the count of participants who had a higher Correct Detection ratio in the second half of the task compared to the first. Ties indicated the count of participants who had no change in performance between the two halves.

*Social Motivation Comparison.* An additional Kruskal-Wallis H Test was performed to explore the effects of social motivation on the vigilance task Correct Detection ratio (H4 – Exp. 2: Correct Detections). Results of the analysis revealed significant differences in the Correct Detection ratio between the groups, $\chi^2(3) = 10.41$, $p = .015$. Post hoc comparisons were made using Dunn's Test with Bonferroni Corrections (see Table 8). Like in the post hoc analysis conducted to compare the Correct Response ratio between groups, results showed a significant difference between Group 2 (ranked leaderboard) and Group 3's (reward incentive) Correct Detection ratio scores, such that Group 2 performed significantly less accurately than Group 3, but no significant differences were observed between any of the other groups. The median and mean rank score for each group were as follows: Group 1 (Control, *Mdn* = 0.683, Mean Rank = 80.97), Group 2 (Leaderboard, *Mdn* = 0.482, Mean Rank = 66.12), Group 3 (Reward Incentive, *Mdn* = 0.750, Mean Rank = 100.43), Group 4 (Combined, *Mdn* = 0.586, Mean Rank = 78.47). Taken together, these findings provide additional partial support for Hypothesis 4 – Exp. 2: Correct Detections, which stated the following: "Significant differences will be seen across the social motivation conditions. Specifically, participants will have the highest accuracy on the sustained attention task in the combined leaderboard/reward incentivization condition, where the most motivating stimuli are present, followed by participants in the standalone gamification leaderboard and reward incentivization conditions, and participants in the control condition will have the lowest accuracy out of any condition.

**Table 8**
*Dunn's Multiple Comparison Test - Correct Detection Ratio*

| Comparison | Group | Test Statistic | Std. Error | Std. Test Statistic | *p* | Adj. *p*[a] |
|---|---|---|---|---|---|---|
| 2-4 | Leaderboard vs. Combined | -12.352 | 10.487 | -1.178 | .239 | 1.000 |
| 2-1 | Leaderboard vs. Control | 14.856 | 11.137 | 1.334 | .182 | 1.000 |
| 2-3 | Leaderboard vs. Reward Incentive | -34.309 | 10.943 | -3.135 | .002 | .010* |
| 4-1 | Combined vs. Control | 2.504 | 10.153 | 0.247 | .805 | 1.000 |
| 4-3 | Combined vs. Reward Incentive | 21.957 | 9.940 | 2.209 | .027 | .163 |
| 1-3 | Control vs. Reward Incentive | -19.453 | 10.624 | -1.831 | .067 | .403 |

a. Bonferroni correction, $*p < .05$

*Full Vigilance Task – Reaction Times.* Of note, unlike Experiment One, no additional outliers were removed in Experiment Two, as no RT participant data fell within the exclusion criteria followed throughout both Experiment One and Experiment Two (>3 *SD*s outside the mean and consistent with the most extreme outliers based on the IQR).

*Experimental vs. Control.* An exploratory Mann Whitney U Test was conducted to examine whether there were RT differences between the control and the combined experimental groups. Results showed no significant differences between the control group (*Mdn* = 639.07, Mean Rank = 75.94) and the combined social motivation group (*Mdn* = 629.46, Mean Rank = 76.02), *U* = 1987.00, *z* = -0.01, *p* = .993.

*First Half vs. Second Half.* A Wilcoxon Sign-Rank Test was conducted to compare RT between the first half of the vigilance task and the second half of the vigilance task. Results showed no significant differences between the first half (*Mdn* = 633.67) and second half of the task (*Mdn* = 637.33), with 63 negative ranks, 68 positive ranks, and 0 ties, *z* = -0.06, *p* = .955.

*Social Motivation Comparison.* A Kruskal-Wallis H Test was conducted to assess RT differences between the social motivation groups. Results indicated no significant differences between the groups (Group 1: Control, *Mdn* = 639.07, Mean Rank = 75.94; Group 2: Ranked Leaderboard, *Mdn* = 637.79, Mean Rank = 85.00; Group 3: Reward Incentive, *Mdn* = 628.84, Mean Rank = 73.65; Group 4: Combined, *Mdn* = 624.81, Mean Rank = 72.56), $\chi^2(3) = 1.64$, *p* = .650.

*Correlation – RT and Accuracy.* An exploratory Spearman correlation was conducted to explore the relationship between RT and vigilance task accuracy. Results showed a significant association between RT and vigilance task accuracy, as measured by the Correct Detection ratio, *r*(149) = -.41, *p* < .001. This finding suggests that higher rates of correct detections were associated with faster average RTs on the vigilance task.

***Boredom***. Gender and ethnicity differences for boredom were evaluated using one-way ANOVAs with Bonferroni post hoc comparisons. No significant gender or ethnicity differences were found for trait or state boredom.

*Full Vigilance Task – Boredom.*

*Experimental vs. Control.* An independent samples t-test was conducted to examine whether there were state boredom differences between the control and the combined experimental groups (H5 – Exp. 2). Results indicated no significant state boredom differences between the control group (*M* = 116.5, *SD* = 36.4) and the combined experimental group (*M* = 115.6, *SD* = 36.1), $t(161) = .13$, $p = .895$, contrary to Hypothesis 5 – Exp. 2, which stated that "Participants who receive the social motivation manipulation, as measured by the combined experimental groups, will have significantly lower state boredom compared to participants who do receive a social motivation manipulation."

*Social Motivation Comparison.* A one-way ANCOVA was conducted to analyze the effect of social motivation on state boredom, while controlling for the effects of trait boredom (H6 - Exp. 2). Before running the analysis, the assumption of homogeneity of regression slopes was evaluated and met, $F(3, 155) = 2.58$, $p = .055$. Additionally, Levene's test for homogeneity of variances was met, $F(3, 159) = .14$, $p = .937$. No significant state boredom differences were seen between the social motivation groups when controlling for trait boredom, $F(3, 158) = 2.38$, $p = .072$, partial $\eta^2 = .04$, contrary to the hypothesis. The adjusted mean for each group was: Group 1 (Control, *M* = 116.3), Group 2 (Leaderboard, *M* = 116.2), Group 3 (Reward Incentive, *M* = 105.7), Group 4 (Combined, *M* = 123.6). The unadjusted mean scores were as follows: Group 1 (Control, *M* = 116.5, *SD* = 36.4), Group 2 (Leaderboard, *M* = 115.7, *SD* = 39.0), Group 3 (Reward Incentive, *M* = 105.8, *SD* = 32.8), Group 4 (Combined, *M* = 123.7, *SD* = 35.2). This finding suggests that after controlling for the impact of trait boredom, the social motivation manipulations had no significant effect on state boredom. However, the covariate, trait boredom, was significantly associated with state boredom $F(1, 158) = 43.83$, $p < .001$, partial $\eta^2 = .22$, suggesting that trait boredom was a significant predictor of state boredom. Overall, this finding did not provide support for Hypothesis 6 – Exp. 2, which stated that "Participants who receive any of the social motivation manipulations will have lower self-reported state boredom scores compared to participants in the control condition."

*Correlations - Trait Boredom, State Boredom, RT, and Accuracy.* A Pearson correlation was conducted to assess the relationship between trait and state boredom. Results revealed a significant positive correlation between the two variables, $r(161) = .46$, $p < .001$. This finding is consistent with what was seen in Experiment One, suggesting that individuals who had high trait boredom levels also tended to report high state boredom levels. Further, exploratory Spearman correlations were conducted to explore the relationships between trait boredom, state boredom, and RT. However, no significant relationships were seen between RT and trait boredom, $r(149) = -.05$, $p = .527$ or between RT and state boredom, $r(149) = -.01$, $p = .931$.

Additionally, Spearman correlations were conducted to explore the relationships between trait boredom, state boredom, and vigilance task accuracy (as measured by the Correct Response ratio and Correct Detection ratio) (H7 – Exp. 2, H8 – Exp. 2). Results showed a significant negative association between trait boredom and the Correct Detection ratio, $r(161) = -.16$, $p = .043$. This finding suggests that as trait boredom levels increased, the number of correct detections to a jump trial decreased. However, no significant association was seen between trait boredom and the Correct Response ratio, contrary to hypotheses, $r(161) = -.11$, $p = .177$. These findings provide partial support for Hypothesis 7 – Exp. 2, which stated that "Participants with higher trait boredom scores will have lower total accuracy scores on the sustained attention task, regardless of the incentives offered."

Significant negative correlations were found between state boredom and the Correct Detection ratio, $r(161) = -.29$, $p < .001$, and between state boredom and Correct Response ratio, $r(161) = -.26$, $p < .001$. These findings indicated that as state boredom levels increased, vigilance task accuracy decreased. Further, these results provide support for Hypothesis 8 – Exp. 2, which stated that “Participants with higher state boredom scores will have lower total accuracy scores on the sustained attention task, regardless of the incentives offered.” Figures 15-17 visually depict the significant negative associations that were found.

**Figure 15**

*Scatterplot Depicting Relationship Between Trait Boredom and Correct Detection Ratio.*

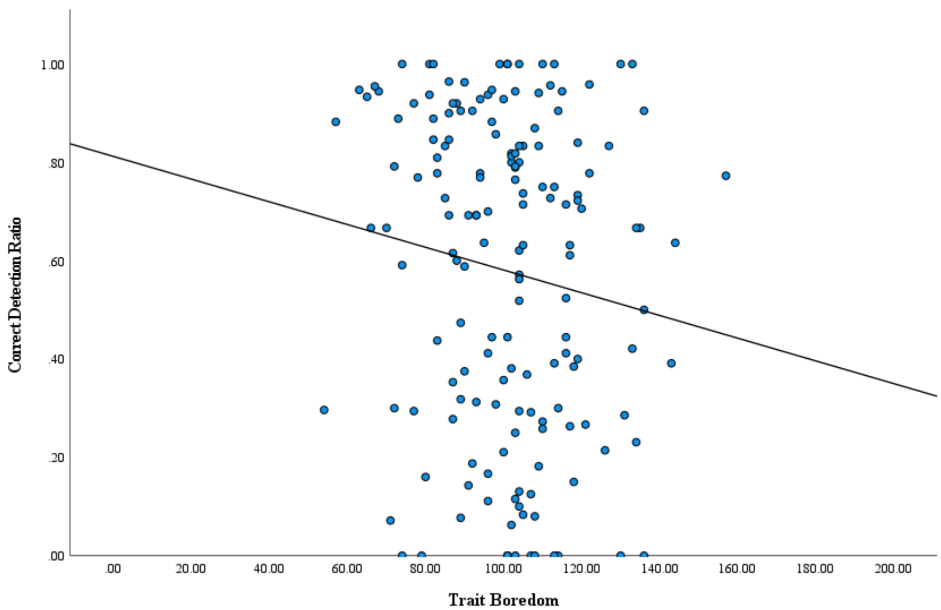


**Figure 16**

*Scatterplot Depicting Relationship Between State Boredom and Correct Response Ratio. Y-axis begins at 0.90 to highlight visual differences.*

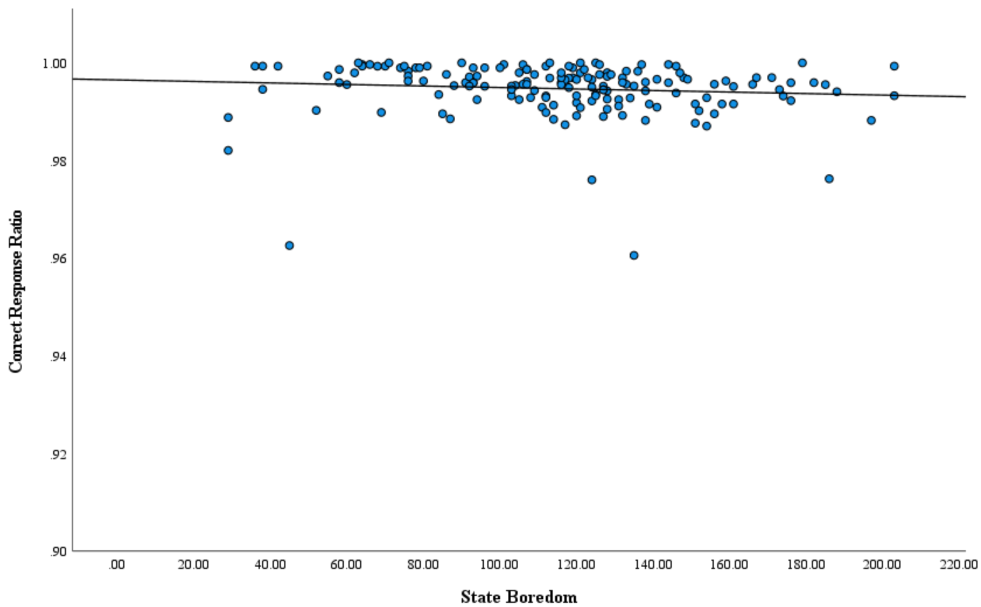
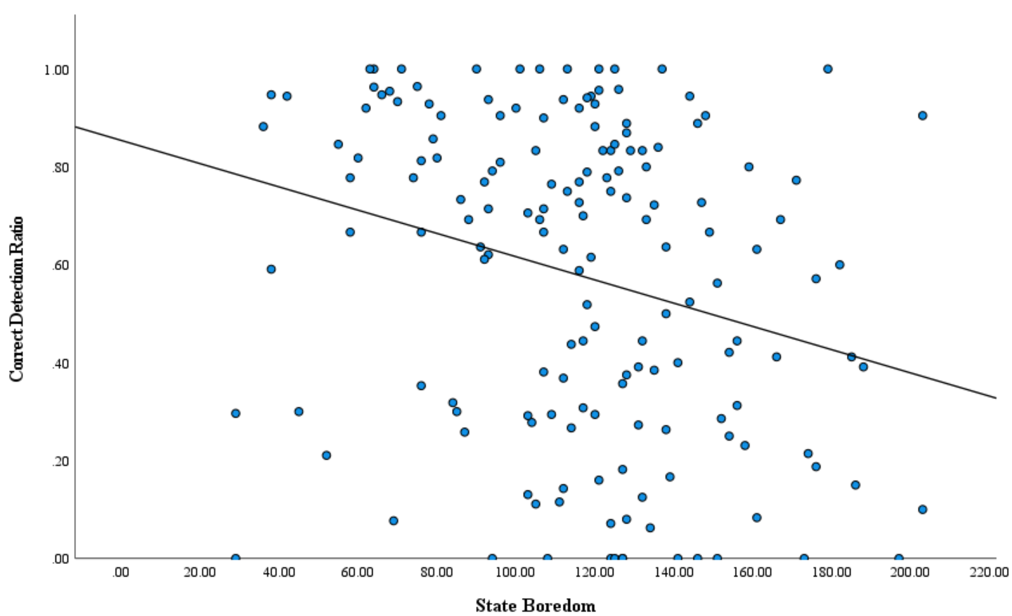



**Figure 17**

*Scatterplot Depicting Relationship Between State Boredom and Correct Detection Ratio.*



## Discussion

Experiment Two sought to investigate the impact of various socially motivating elements on sustained attention performance, as measured by overall accuracy and RT. Specifically, the impact of different competition presentations on performance during a sustained attention task were examined. The effects of gamification on vigilance task accuracy and state boredom were explored using a high score ranked leaderboard that displayed other participants' scores on the same task. Additionally, reward incentivization via the offering of an additional SONA point if participants scored a high enough score was tested as an alternative format. Of note, both types of competition formats explored in the current study were examples of *extrinsically* motivating stimuli, as they both sought to achieve a goal because of the influence of an external factor that led to a separable outcome (Ryan & Deci, 2000). Overall, the findings suggested that within the context of the current sample, participants who were shown a reward incentive message displayed higher accuracy on the vigilance task than participants who viewed a ranked leaderboard. However, neither group had significantly different accuracy scores compared to the control group. Additionally, the socially motivating stimuli were initially effective in increasing accuracy within the first ten minutes of the task, but they did not impact performance compared to the control when compared across the full task. Table 9 lists all hypotheses from Experiment Two and indicates whether support was found in the current study. Further, the following paragraphs outline the key findings of Experiment Two.

**Table 9**

*Hypotheses - Experiment Two*

| Hypothesis # | Hypothesis | Support (Yes/No) |
|---|---|---|
| H1 | Performance accuracy on the sustained attention task will improve within subjects when presented with a social motivation manipulation compared to performance on a training block where no manipulation is present. | Yes |
| H2 | Participants who receive a social motivation manipulation will perform significantly more accurately on the task compared to participants who are in the control condition. | No |
| H3 | Performance accuracy on the sustained attention task will decrease within subjects from the first half of the task compared to the second half of the task. | Partial |
| H4 | Significant differences will be seen across the social motivation conditions. Specifically, participants will have the highest accuracy on the sustained attention task in the combined leaderboard/reward incentivization condition, where the most motivating stimuli are present, followed by participants in the standalone gamification leaderboard and reward incentivization conditions, and participants in the control condition will have the lowest accuracy out of any condition. | Partial |
| H5 | Participants who receive the social motivation manipulation, as measured by the combined experimental groups, will have significantly lower state boredom compared to participants who do receive a social motivation manipulation. | No |
| H6 | Participants who receive any of the three standalone social motivation manipulations will have lower self-reported state boredom scores compared to participants in the control condition. | No |
| H7 | Participants with higher trait boredom scores will have lower total accuracy scores on the sustained attention task, regardless of the incentives offered. | Partial |
| H8 | Participants with higher state boredom scores will have lower total accuracy scores on the sustained attention task, regardless of the incentives offered. | Yes |

When comparing performance accuracy across the social motivation groups, participants who received a reward incentivization offer prior to completing the vigilance task displayed significantly higher accuracy scores than participants who were shown the ranked leaderboard. This finding is akin to previous research that found that offering high financial rewards for performance resulted in prolonged and stronger attention compared to performing the same task when no obvious financial reward was offered (Begleiter et al., 1983; Hömberg et al., 1981; Oken, 2010). However, no significant accuracy differences were seen across any of the other groups in the current study, including the control group. It is possible that the leaderboard was ineffective, resulting in reduced performance, while the reward incentive was effective at increasing performance, explaining why no significant differences were seen compared to the control group. Further, no significant RT differences were seen across any of the groups.

An additional component of Experiment Two sought to explore the impact of the various competition types on state boredom level, as state boredom is associated with performance decrements on vigilance-based tasks (Hunter & Eastwood, 2018; Petranker & Eastwood, 2021), and introducing factors to reduce the effects of boredom may be beneficial. This effect was evaluated while controlling for the impact of trait boredom, as trait boredom is one's predisposition to feeling bored and is more stable across environments than state boredom (Bench & Lench, 2019; Fahlman et al., 2013; Farmer & Sundberg, 1986). However, contrary to hypotheses, no significant state boredom differences were found between the social motivation groups. Previous research suggests associations between competition and reductions in experiences of boredom in a variety of contexts, such as with athletes and sports engagement (Velasco & Jorda, 2020) and employees in work settings (Dishon-Berkovits et al., 2023). However, in the context of the current study, it appears that competition was not effective at reducing state boredom in the context of a sustained attention task, where tasks are often cognitively demanding (Davies & Parasuraman, 1982; Hancock & Warm, 1989; Hockey, 1997; Kahneman, 1973; Parasuraman et al., 1987; Wickens, 2002), monotonous, and repetitive (Manly et al., 1999; Robertson et al., 1997). Additionally, it is also a possibility that in this study, the socially motivating stimuli were not strong enough to elicit changes in self-reported experiences of state boredom across the sample.

Similarly, no performance differences (RT, state boredom, accuracy) were found when comparing the combined “social motivation” group (e.g., all participants who received any type of competition) to the control group. It is possible that in the current study, the motivating message was not mitigating enough to engage attention and reduce self-reported state boredom on the vigilance task, which was designed to be monotonous and thus can be boring (Manly et al., 1999; Robertson et al., 1997). These findings do not support the current study’s hypotheses, where it was anticipated that participants who received a competitive stimulus would perform more accurately and experience less boredom than participants in the control group. However, the findings may provide insight about the efficacy of the competition formats, where participants in the current study performed more accurately when offered an additional SONA credit for accurate performance compared to participants who were shown a “gamified” version of the task, where a ranked leaderboard with high scores was presented. In fact, participants in the SONA incentive group had the highest accuracy across all the social motivation groups, including the combined group (where both the ranked leaderboard and SONA incentive message were presented). This finding was unexpected, where the hypothesis was that participants in the combined condition would perform the most accurately as they received the most motivating stimuli. However, it is possible that the ranked leaderboard in this study was ineffective and thus may have negatively impacted the efficacy of the reward incentivization in the combined condition. In the current study, participants may have felt a stronger sense of motivation when offered a reward for performance as compared to in the ranked leaderboard condition, that did not offer any type of compensation for performance. This sense of drive and motivation could possibly explain why participants performed more accurately in the reward incentive condition compared to the ranked leaderboard condition.

In addition to measuring task performance differences across social motivation groups, the findings of the current study suggested that there were within-subjects accuracy differences pre-manipulation versus post-manipulation. Rather, accuracy in the ten-minute training block was compared to accuracy in the first ten minutes of the actual task in all experimental groups (not control group) to test the impact of the socially motivating stimuli. Results showed that participants performed more accurately, as measured by the Correct Response ratio, on the vigilance task after they received a socially motivating stimulus compared to in the training block. This finding may provide evidence that introducing a competitive element was initially effective in motivating participants, explaining why participants had higher performance accuracy scores in the first ten minutes of the full task compared to in the training. Previous research has demonstrated evidence that various socially motivating stimuli, such as introducing an incentivization (Garner et al., 2023), and "points-based rewards" (Robison & Nguyen, 2023, pp. 1256) are effective tools to reduce the extent of the vigilance decrement. This finding in the current study supports this research, providing evidence that participants had higher accuracy in the first ten minutes after receiving the social motivation manipulation compared to in the training. However, it is also possible that observed changes were influenced by practice effects, where the participant became familiar with the task over time. To explore this theory further, an additional exploratory analysis was conducted to test for performance differences in the control group, who did not receive the message manipulation. Results in this supplementary analysis indicated a significant difference, where participants had increased accuracy in the first ten minutes compared to the control. Therefore, it is likely that practice effects influenced the observed differences in the training analysis to some extent.

Further, a significant vigilance decrement was observed when comparing performance accuracy between the first half of the task and the second half of the task, where participants correctly detected significantly fewer jump trials in the second half than in the first. This finding supports the study's hypotheses and is consistent with well-established evidence of the vigilance decrement, where errors are seen over time due to the cognitive demands of the task (Hancock & Warm, 1989; Hockey, 1997; Parasuraman & Moulouna, 1987). However, no vigilance decrement was observed between the two halves when accuracy was measured by the Correct Response ratio, where correct performance also recognized trials where the "correct" response was to press nothing on the keyboard. The Correct Detection ratio focused on instances that required a physical response from the participant, so it may be that these trials were more cognitively demanding than what was captured by the Correct Response ratio, explaining why a vigilance decrement was observed in the former but not the latter. Additionally, no vigilance decrement was observed between the first and second halves of the tasks when performance was measured by RT, a metric that is often considered an indicator of attention on vigilance tasks (Yamashita et al., 2021).

The relationship between trait boredom, state boredom, and overall vigilance task accuracy were also explored. For the association between state boredom and task accuracy, results were consistent with what was anticipated in the hypotheses. Specifically, higher self-reported state boredom levels were associated with lower accuracy on the vigilance task, as measured by both the Correct Response ratio and the Correct Detection ratio. A person who is uninterested or uncaring towards a task is less likely to be as vigilant in their attention and performance compared to individuals who have a strong sense of motivation or care (Oken, 2010). In the context of sustained attention tasks, it appears that high in-the-moment feelings of boredom also relate to decreased vigilance. A similar association was found for the relationship between trait boredom and the Correct Detection ratio, the ratio of instances when participants correctly responded to a jump trial on the task. These findings are consistent with Mindlessness Theory that states vigilance decrements are due to the monotony of the task. According to the theory, this monotony results in higher experiences of boredom and mind-wandering, eventually contributing to an increased number of errors (Manly et al., 1999; Robertson et al., 1997). Additionally, these findings provide additional support for previous research that shows that high trait boredom (Hunter & Eastwood, 2018; Malkovsky et al., 2012) and state boredom (Hunter & Eastwood, 2018; Petranker & Eastwood, 2021) are associated with lower vigilance task accuracy. Of note, in the current study, no significant relationship was found between trait boredom and the Correct Response ratio, the ratio of total "correct" responses based on the given trial. This finding is not consistent with previous research and warrants additional research and consideration to determine the lack of significant findings in the current study. Overall, this study demonstrated the importance of considering both trait and state boredom when evaluating performance on vigilance tasks, as high experiences of boredom may lead to an uptick in errors on tasks that require high vigilance capabilities over time. RT's association with correct detection performance accuracy and boredom was also examined in exploratory analyses. Results indicated no relationship between trait boredom, state boredom, and RT. However, a significant association was found between RT and accuracy, such that faster RTs were associated with more correct detections. This finding supports research conducted by Morais et al. (2024), who found that faster RTs were correlated with a greater number of correct detections.

## General Discussion

Together, these two complementary experiments explored the effectiveness of various factors in improving performance on vigilance-based tasks. Experiment One investigated the effects of varying signal words on vigilance performance and boredom, and Experiment Two assessed the impact of varying types of social motivation on vigilance performance and boredom. Taken together, both studies examined several techniques to improve performance on sustained attention tasks. These findings sought to provide evidence of strategies to reduce errors on sustained attention tasks, with the goal being to identify strategies that can be implemented into protocols to improve safety and efficiency in a variety of domains.

Experiment One found evidence that the presentation of a combined motivating message/signal word/alert symbol warning of consequences for noncompliance resulted in increased performance accuracy and faster RTs on a vigilance task compared to participants who received no message prior to completing the same task. Specifically, participants who received the message "[SIGNAL WORD]: If you do not complete the following task with at least **95%** accuracy, you will be required to re-start the activity" in conjunction with a signal word and alert symbol performed more accurately on a 50-minute trial of the Mackworth Clock Test compared to participants who completed the same activity but were presented with no message. This finding is consistent with previous research that has demonstrated that presenting a motivating stimulus prior to a sustained attention task can increase performance accuracy (McGough & Mayhorn, 2022). Further, research has demonstrated that incentivization increases accuracy on sustained attention tasks (Garner et al., 2024), and RT is a metric often considered to be an indicator of attention (Yamashita et al., 2021). Therefore, this finding provides evidence that utilizing motivational messages in conjunction with signal words and alert symbols may be an effective strategy to increase accuracy on vigilance-based tasks. Further, when comparing correct detection accuracy differences between the signal word/text highlight color variations that preceded the motivational message, the control group performed significantly less accurately on the task than several of the experimental groups. However, no significant accuracy differences were observed between any of the signal word/text highlight color groups themselves, contrary to hypotheses. These findings indicate that varying the signal word/text highlight color itself was ineffective in altering performance. Notably, no significant accuracy differences were seen between any of the groups when accuracy was measured by the Correct Response ratio. This discrepancy may suggest that the presence of a motivating warning message is more effective in improving the accuracy of target detection rather than reducing incorrect detections and missed detections, but additional research is needed to explore this hypothesis. Further, no significant state boredom differences were seen between the signal word/text highlight color groups, contrary to hypotheses. However, trait boredom was found to be a significant predictor of state boredom, which is consistent with previous literature that demonstrates an association between trait and state boredom (Hunter & Eastwood, 2018). Moreover, negative associations between trait/state boredom and vigilance task accuracy were observed such that high trait/state boredom levels were related to lower accuracy on the Mackworth Clock Test. These findings provide additional support for previous literature, where boredom as an individual difference has been associated with reduced performance on vigilance tasks (Hunter & Eastwood, 2018; Malkovsky et al., 2012; Robertson, 1997). Additionally, negative associations were found between RT and correct detections, providing additional support that faster RTs are associated with increased accuracy on vigilance tasks (Morais et al., 2024).

Experiment Two explored the impact of presenting varying types of socially motivating competition formats when participants engaged with the vigilance task. Four social motivation groups were considered: a control condition (where participants received no stimulus prior to completing the task), a gamification ranked leaderboard condition, a reward incentivization condition, and a combined condition (where participants received the leaderboard and reward incentive). Results of the study failed to provide evidence of significant accuracy differences between participants who received a socially motivating stimulus compared to participants who received no stimulus before the task. However, significant accuracy differences were seen between the ranked leaderboard and reward incentivization conditions, where participants who received the reward incentive had higher accuracy scores on the task than participants who were shown the ranked leaderboard before beginning the task. Additionally, no significant self-reported state boredom differences were found between the social motivation groups after completing the vigilance task, contrary to hypotheses but consistent with the findings of Experiment One. This difference was explored while controlling for the effect of trait boredom, as trait boredom represents a person's predisposition to feeling boredom and is more stable across environments than state boredom (Bench & Lench, 2019; Fahlman et al., 2013; Farmer & Sundberg, 1986). Previous research has demonstrated that providing a motivating stimulus can influence performance on tasks (Ryan & Deci, 2000; Wogalter, 2006b). Additionally, associations have been observed specifically between competition and reduced boredom in varying contexts (Dishon-Berkovits et al., 2023; Velasco & Jorda, 2020). However, competition, which broadly extends to motivational stimuli in general, can also hinder performance on tasks and change one's approach to completing the task, depending on how it is perceived. Rather, how competition is presented can impact perception of the stimulus either positively or negatively, as explained by Reeve (2023), who described the differences that can make competition either constructive or destructive. In the current studies, individual differences in perception to the motivational stimuli may have resulted in the lack of significant boredom differences between groups. Additionally, it is also a possibility that in this study, the message manipulation in Experiment One and the socially motivating stimuli in Experiment Two were not strong enough to elicit changes in self-reported experiences of state boredom across the sample. In both studies, however, trait boredom was identified as a significant predictor of state boredom, consistent with previous boredom literature that shows associations between the two variables (Hunter & Eastwood, 2018). Additionally, as in Experiment One, Experiment Two results indicated negative associations between RT and correct detections, providing additional support that faster RTs are associated with increased accuracy on vigilance tasks (Morais et al., 2024).

Results from both experiments also provided evidence of the relationship between trait/state boredom and vigilance task accuracy, where high boredom levels were related to lower accuracy on the vigilance task. Interestingly, Experiment Two varied from Experiment One in the relationship between trait boredom and task accuracy when task accuracy was measured by overall correct responses, where no significant relationship was seen in Experiment Two. However, trait boredom and task accuracy were associated with one another when task accuracy was measured by overall correct detections, consistent with the findings of Experiment One.

In both experiments, significant accuracy differences, as measured by overall correct responses, were observed between the training block (pre-manipulation) and the first ten minutes of the full task. The goal of including the experimental manipulations across both experiments (Exp.1: signal word variation, Exp. 2: social motivation variation) and the motivational warning message itself (Exp. 1) was to capture the participant's attention. In the case of the signal words, they were intended to alert participants of the potential for an undesirable outcome (ANSI, 2002; Mayhorn et al., 2015). Additionally, for both experiments, the presence of the warning message and the social motivation variations were intended to be types of extrinsic motivators, where there was an outside factor influencing the user's performance on the task (Ryan & Deci, 1985). Taken together, the results from both experiments provide evidence to support the effectiveness of extrinsic motivators in encouraging performance accuracy on vigilance tasks, such that the combined motivational message/signal word/alert symbol (Exp. 1) and the socially motivating competition methods (Exp. 2) were impactful in capturing participants' attention within the first ten minutes of their presentation, as depicted by a higher accuracy score after the manipulation was introduced compared to in the training. However, it is likely that observed changes were influenced by or due to practice effects to some extent, where the participant became familiar with the task from the training to the first 10 minutes of the full task.

When assessing for the presence of the vigilance decrement, regardless of manipulations, a significant accuracy decrement was seen between the first and second halves of Experiment Two, only when accuracy was measured by overall correct detections (correctly responded to a jump trial). In this instance, participants performed significantly less accurately in the second half of the task compared to the first half, consistent with the strong evidence of the vigilance decrement across many other studies (Davies & Parasuraman, 1982; Hancock & Warm, 1989; Hockey, 1997; Kahneman, 1973; Manly et al., 1999; Parasuraman et al., 1987; Robertson et al., 1997; Wickens, 2002). However, no evidence of the vigilance decrement was observed in

Experiment One or in Experiment Two when accuracy was measured by overall correct responses (e.g., correctly withheld response on normal trials, did not miss jump trials). These results suggest that vigilance is likely impacted by many other cognitive processes other than motivation. Other task factors that can impact vigilance on sustained attention tasks are the duration of the activity, the intensity of the task, and the target rate probability (Parasuraman, 1998b). These variables could explain why no significant decrements were seen between the first half and second halves of the task in Experiment One or when accuracy was measured by correct responses in Experiment Two. The length of the task (50 minutes) and the small amount of jump trials (slightly more than half a percent jump rate) may have impacted vigilance performance regardless of the presentation of any overt motivational stimuli. Further, the lack of vigilance decrement in both studies when accuracy was measured by correct responses may be explained by the method that the Correct Response ratio was calculated versus the Correct Detection ratio. Accuracy when measured by the correct responses is inherently less cognitively demanding than accuracy as measured by correct detections because the Correct Detection ratio focused on instances that required a physical response from the participant, while the Correct Response ratio included instances that required no response. Therefore, it may be that the Correct Detection ratio captured a more cognitively demanding set of trials than what was captured by the Correct Response ratio, explaining why a vigilance decrement was observed when accuracy was measured by correct detections (in Experiment Two) but not correct responses. In this case, less resource-demanding activities would elicit less of a vigilance decrement over time, consistent with Resource Theory. (Davies & Parasuraman, 1982; Hancock & Warm, 1989; Hockey, 1997; Kahneman, 1973; Parasuraman et al., 1987; Wickens, 2002).

## Limitations

Several limitations existed in the execution of both experiments. In Experiment One, the study's design where signal word and text highlight color were treated as a combined variable prevented the ability to test for interaction effects. This design choice was made intentionally, as the text highlight color variable on its own lacked meaning for the purposes of this research when not paired with the signal word. For example, testing the effect of the color orange without any signal word lacks meaning for the recipient and thus was not explored. Similarly, Experiment One's design did not allow the researchers to test the individual effect of the motivational message without any signal word present. While results showed evidence for the effectiveness of the motivational message when it was presented with at least one variation of all the signal words, there was no condition that tested the impact of the motivational message on its own without any signal word. Therefore, the study's design prevented the individual assessment of each component (motivational message/signal word/alert symbol) to understand their specific contribution.

For both experiments, the entire sample consisted of Introduction to Psychology students participating in research as part of a course requirement. However, in Experiment Two, this created a limitation. It is possible that this factor influenced participant behavior and performance. Participants were required to earn a certain number of SONA credits as part of their course requirement, and that likely added additional influence on the reward incentivization condition, where participants were offered one additional SONA credit for accurate performance on the vigilance task. This factor may have biased performance on the task within the context of the current sample, explaining why participants in the reward incentive condition performed significantly more accurately than participants in the leaderboard condition. Further, the

Mackworth Clock Test, the vigilance task used in both experiments, is a laboratory-based task. It is a basic paradigm that may not be fully representative of real-world vigilance tasks. Additionally, a limitation for both experiments was that aside from boredom, there are many other individual and environmental differences that could impact sustained attention performance. For example, individual differences in competitiveness may impact the effectiveness of using competition as a socially motivating stimulus. The current studies' designs did not control for other individual differences aside from the effect of trait boredom.

### Future Directions

Future research should further explore the individual impact of the motivational message from Experiment One with and without the presence of signal words to get a better understanding of the impact of the message itself versus the message alongside any type of signal word variations. This would provide insight into whether the motivational message is an effective tool to improve performance on vigilance tasks when it is presented without the signal word, or if the presence of the signal word in some way influences the perception of the message.

Additionally, more research is needed to examine the impact of individual and environmental differences other than trait boredom that could influence vigilance and performance on sustained attention tasks. For example, research has shown that sustained attention can be impacted by various disorders such as attention deficit hyperactivity disorder (ADHD) (Tucha et al., 2015) and major depressive disorder (Piani et al., 2022). Other research suggests that sustained attention performance changes with age. For instance, in a 10,430-person sample conducted by Fortenbaugh et al. (2015), sustained attention ability was found to peak in the early 40s and slowly decline in older individuals. These factors may be important to consider for high stakes jobs that require strong vigilance capabilities, and research is needed to explore whether any mitigation strategies can be implemented to reduce their impact on performance.

To increase the likelihood of finding generalizable results, this study should be replicated across different samples to explore the impact of the socially motivating competition methods, with a specific focus on testing the impact of reward incentivization in samples without the same biases that were in the current studies. For instance, research could be conducted to test a small monetary reward incentivization instead of SONA research credit, that was a course requirement for all participants in the current sample. Additionally, the effects of signal words and social motivation on sustained attention performance should be explored in more ecologically valid domains, as the Mackworth Clock Test may not be representative of real-world vigilance tasks.

## Conclusion

These findings provide partial support that utilizing various types of extrinsic motivation can promote accuracy on vigilance-based tasks. Findings were mixed between Experiment One and Experiment Two, but overall findings suggest that some types of motivators/competition methods may be helpful in reducing errors on sustained attention tasks. However, no evidence was found that would support the notion that varying signal words would change a user's performance on the task based on the signal word itself. While it appears that the combined impact of a motivational message in conjunction with a signal word and alert symbol influenced performance outcomes compared to when no message was shown before the task, varying the signal word itself did not alter performance. Additionally, in the context of the current studies, participants who received the message offering a reward incentive displayed higher accuracy on a sustained attention task compared to participants who viewed a "gamified" high score ranked leaderboard. Of note, no significant accuracy differences were observed between the control condition and any of the competition formats over the course of the entire duration of the task, but participants in the social motivation groups performed more accurately in the first ten minutes of the task compared to in the training block. Interestingly, results did not suggest that any type of motivating stimulus across either experiment influenced state boredom levels after the completion of the vigilance task. However, evidence was garnered for the relationship between boredom and accuracy on vigilance tasks, such that high boredom is associated with reduced accuracy on tasks that require high levels of vigilance over time. Further, support was found for associations between correct detections and RT, where a higher number of correct detections on the vigilance task were related to faster RTs. Taken together, these findings may inform design considerations for protocols and human-centered technologies that require strong vigilance capabilities from the user. Incorporating motivational elements as appropriate may be helpful in promoting performance accuracy in certain domains. For example, incorporating language that outlines the risk of non-compliance at the start of a task or providing extrinsic goals/milestones to work towards in the form of motivating messaging may be effective in promoting performance accuracy on workplace activities requiring sustained attention. Further, the impact of boredom should be considered when designing protocols that heavily rely on sustained attention. For instance, it may be valuable for employers to seek feedback from employees regarding methods to mitigate boredom when completing sustained attention tasks that are particularly repetitive, as boredom may affect performance outcomes and be associated with an increase in errors.

## REFERENCES


Abrams, D., & Hogg, M. A. (Eds.). (1990). *Social identity theory: Constructive and critical advances.* Springer-Verlag Publishing.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behaviour and Human Decision Processes, 50*(2), 179-211. https://doi.org/10.1016/0749-5978(91)90020-T

Allan Cheyne, J., Solman, G. J., Carriere, J. S. & Smilek, D. Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition 111*, 98–113. https://doi.org/10.1016/j.cognition.2008.12.009

Ambach, W., Assmann, B., Wielandt, B., & Vaitl, D. (2019). Evaluative observation in a concealed information test. *Frontiers in Psychiatry, 10*, 636. https://doi.org/10.3389/fpsyt.2019.00636

ANSI. (2002). *Accredited standards committee on safety signs and colors. Z535.1-5.* National Electrical Manufacturers Association.

Begleiter H., Porjesz B., Chou C.L., Aunon J.I. (1983). P3 and stimulus incentive values. *Psychophysiology*, *20*(1), 95–101. https://doi.org/10.1111/j.1469-8986.1983.tb00909.x

Bench, S. W., & Lench, H. C. (2019). Boredom as a seeking state: Boredom prompts the pursuit of novel (even negative) experiences. *Emotion, 19*(2), 242–254. https://doi.org/10.1037/emo0000433

Bond, C. F., & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin, 94*, 265-292. https://doi.org/10.1037/0033-2909.94.2.265

Chapanis, A. 1994. Hazards associated with three signal words and four colours on warning signs. *Ergonomics, 37*, 265-275. https://doi.org/10.1080/00140139408963644

Conzola, V. C., & Wogalter, M. S. (2001). A communication-human information processing (C-HIP) approach to warning effectiveness in the workplace. *Journal of Risk Research*, *4*(4), 309–322. https://doi.org/10.1080/13669870110062712

Cottrell, N. B. (1972). Social facilitation. In C. G. McClintock (Ed.), *Experimental social psychology*. Holt, Rinehart & Winston.

Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance*. Academic Press.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* Springer. https://doi.org/10.1007/978-1-4899-2271-7

DiMenichi, B.C., & Tricomi, E. (2015). The power of competition: Effects of social motivation on attention, sustained physical effort, and learning. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01282

Dishon-Berkovits, M., Bakker, A. B., & Peters, P. (2023). Playful work design, engagement and performance: the moderating roles of boredom and conscientiousness. *The International Journal of Human Resource Management*, *35*(2), 256–283. https://doi.org/10.1080/09585192.2023.2227920

Fahlman, S. A., Mercer-Lynn, K. B., Flora, D. B., & Eastwood, J.D. (2013). Development and Validation of the Multidimensional State Boredom Scale. *Assessment, 20*(1), 68–85. https://doi.org/10.1177/1073191111421303

Farmer, R., & Sundberg, N. D. (1986). Boredom proneness: The development and correlates of a new scale. *Journal of Personality Assessment*, *50*(1), 4–17. https://doi.org/10.1207/s15327752jpa5001_2

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

FMC Corporation. (1985). *Product safety sign and label system.* FMC Corporation.

Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., & Esterman, M. (2015). Sustained attention across the life span in a sample of 10,000: Dissociating ability and strategy. *Psychological science*, *26*(9), 1497–1510. https://doi.org/10.1177/0956797615594896

Garner, L. D., Shuman, J., & Robison, M. K. (2024). Piece-rate time-based incentives improve sustained attention. *Journal of Applied Research in Memory and Cognition, 13*(3), 342–353. https://doi.org/10.1037/mac0000141

Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinsky, T. L., Szalma, J. L., & Parasuraman, R. (2003). The vigilance decrement reflects limitations in effortful attention, not mindlessness. *Human Factors*, *45*(3), 349– 359. https://doi.org/10.1518/hfes.45.3.349.27253

Guerin, B. (2010). *Social facilitation.* Wiley.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, *31*(5), 519–537. https://doi.org/10.1177/001872088903100503

Hockey, R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*(1-3), 73-93. https://doi.org/10.1016/S0301-0511(96)05223-4

Hömberg V, Grnewald G, Grnewald-Zuberbier E. (1981). The variation of p300 amplitude in a money winning paradigm in children. *Psychophysiology, 18*(3), 258–262. https://doi.org/10.1111/j.1469-8986.1981.tb03030.x

Hunter, A., & Eastwood, J.D. (2018). Does state boredom cause failures of attention? Examining the relations between trait boredom, state boredom, and sustained attention. *Experimental Brain Research, 236*(9). https://doi.org/10.1007/S00221-016-4749-7

Ishihara, S. (1917). *Tests for colour-blindness*. Hongo Harukicho.

Jerison, H.J. (1970). Vigilance: A paradigm and some physiological speculations. *Acta Psychologica, 33*, 367-380. https://doi.org/10.1016/0001-6918(70)90147-2

Ko, L., Komarov, O., Hairston, D.W., Jung, T., Lin, C. (2017). Sustained attention in real classroom settings: An EEG study. *Frontiers in Human Science, 11*, 388. https://doi.org/10.3389/fnhum.2017.00388

Kahneman, D. (1973). *Attention and effort* (Vol. 1063, pp. 218-226). Prentice-Hall.

Klehe, U. C., Anderson, N., & Hoefnagels, E. A. (2007). Social facilitation and inhibition during maximum versus typical performance situations. *Human Performance, 20*, 223-239. https://doi.org/10.1080/08959280701333040

Latané, B. (1981). The psychology of social impact. *American Psychologist, 36*(4), 343–356. https://doi.org/10.1037/0003-066X.36.4.343

Lichstein, K. L., Riedel, B. W., & Richman, S. L. (2000). The Mackworth Clock Test: a computerized version. *The Journal of Psychology*, *134*(2), 153–161. https://doi.org/10.1080/00223980009600858

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470214808416738

Malkovsky E, Merrifeld C, Goldberg Y, & Danckert J (2012). Exploring the relationship between boredom and sustained attention. *Experimental Brain Research*, *221*(1), 59–67. https://doi.org/10.1007/s00221-012-3147-z

Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent mind: Further investigations of sustained attention to response. *Neuropsychologia*, *37*(6), 661–670. https://doi.org/10.1016/s0028-3932(98)00127-4

Markus, H. (1978). The effect of mere presence on social facilitation: An unobtrusive test. *Journal of Experimental Social Psychology*, *14*(4), 389–397. https://doi.org/10.1016/0022-1031(78)90034-3

Mayhorn, C. B., Wogalter, M. S., & Laughery, K. R. (2015). Analysis and design of warnings in the workplace. In J. R. Wilson & S. Sharples (Eds.). *Evaluation of human work* (pp. 331-358). Taylor & Francis.

Mayhorn, C. B., Wogalter, M. S., Bell, J. L., & Shaver, E. F. (2004). What does code red mean? *Ergonomics in Des*ign, *12*(4), 12-14. https://doi.org/10.1177/106480460401200404

McGough, O. H., & Mayhorn, C. B. (2022). In response to warnings: Exploring individual differences in sustained attention performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 231–235. https://doi.org/10.1177/1071181322661095

McGough, O. H., & Mayhorn, C. B. (2023). Exploring strategies to improve performance accuracy on vigilance-based tasks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *67*(1), 1116-1121. https://doi.org/10.1177/21695067231192440

Morais, G. L. d., Pinho, T. O. R., Crespim, L., & Pinto Neto, O. (2024). Quantifying the transition from unconscious to conscious detection of temporal patterns in vigilance tasks: A unique adaptation of Mackworth's Clock Test. *Neurology International*, *16*(5), 945-957. https://doi.org/10.3390/neurolint16050071

Oken, B. S., Salinsky, M. C., & Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *117*(9), 1885–1901. https://doi.org/10.1016/j.clinph.2006.01.017

Parasuraman, R., & Mouloua, M. Interaction of signal discriminability and task type in vigilance decrement. (1987). *Perception & Psychophysics, 41*, 17–22. https://doi.org/10.3758/BF03208208

Parasuraman, R., Warm, J. S., & See, J. E. (1998b). Brain systems of vigilance. In R. Parasuraman (Ed.), *The attentive brain* (pp. 221–256). The MIT Press.

Peckham, G. M. (2006). ISO design standards for safety signs and label s. In M. S. Wogalter (Ed.), *Handbook of warnings* (pp. 455-462). Lawrence Erlbaum Associates.

Petranker R., Eastwood, J.D. (2021). Sitting with it: An investigation of the relationship between trait mindfulness and sustained attention. *Consciousness and Cognition, 90*. https://doi.org/10.1016/j.concog.2021.103101

Piani, M. C., Maggioni, E., Delvecchio, G., & Brambilla, P. (2022). Sustained attention alterations in major depressive disorder: A review of fMRI studies employing Go/No-Go and CPT tasks. *Journal of Affective Disorders, 303*, 98-113. https://doi.org/10.1016/j.jad.2022.02.003

Platania, J., & Moran, G. P. (2001). Social facilitation as a function of the mere presence of others. *The Journal of Social Psychology*, *141*(2), 190–197. https://doi.org/10.1080/00224540109600546

Reeve, J. (2023). Competition can enhance motivation—but typically undermines it. In M. Bong, J. Reeve, & S. Kim (Eds.), *Motivation science: Controversies and insights* (online ed.). Oxford Academic. https://doi.org/10.1093/oso/9780197662359.003.0028

Robertson, I.H., Manly, T., Andrade, J., Baddeley, B.T., Yiend, J. (1997). 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758. https://doi.org/10.1016/s0028-3932(97)00015-8

Robison, M. K., & Nguyen, B. (2023). Competition and reward structures nearly eliminate time-on-task performance decrements: Implications for theories of vigilance and mental effort. *Journal of Experimental Psychology: Human Perception and Performance*, *49*(9), 1256–1270. https://doi.org/10.1037/xhp0001148

Rogers, W. A., Lamson, N., & Rousseau, G. K. (2000). Warning research: An integrative perspective. *Human Factors*, *42*(1), 102-139. https://doi.org/10.1518/001872000779656624

Ryan, R.M., Deci, E.L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 64-67. https://doi.org/10.1006/ceps.1999.1020

Sanna, L. J., & Shotland, R. L. (1990). Valence of anticipated evaluation and social facilitation. *Journal of Experimental Social Psychology, 26*(1), 82–92.

https://doi.org/10.1016/0022-1031(90)90063-R

Schmitt, B. H., Gilovich, T., Goore, N., & Joseph, L. (1986). Mere presence and social facilitation: One more time. *Journal of Experimental Social Psychology, 22*(3), 242–248. https://doi.org/10.1016/0022-1031(86)90027-2

See, J. E., Howe, S. R., Warm, J. S., & Dember, W. N. (1995). Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin, 117*(2), 230–249. https://doi.org/10.1037/0033-2909.117.2.230

Shaw, T.H., Matthews, G., Warm, J.S., Finomore, V.S., Silverman, L., Costa, P.T. (2010). Individual differences in vigilance: Personality, ability, and states of stress. *Journal of Research in Personality, 44*(3), 297-308. https://doi.org/10.1016/j.jrp.2010.02.007

Smallwood, J., Davies, J.B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition, 13*(4), 657–690. https://doi.org/10.1016/j.concog.2004.06.003

Smit, A. S., Eling, P. A. T. M., & Coenen, A. M. L. (2004). Mental effort causes vigilance decrease due to resource depletion. *Acta Psychologica*, *115*(1), 35–42. https://doi.org/10.1016/j.actpsy.2003.11.001

Sona Systems. (2024). *Sona Systems* [Online participant management system]. https://www.sona-systems.com

Steinmetz, J., & Pfattheicher, S. (2017). Beyond social facilitation: A review of the far-reaching effects of social attention. *Social Cognition, 35*(5), 585-599. https://doi.org/10.1521/soco.2017.35.5.585

Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096-1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction time experiments. *Teaching of Psychology, 44*(1), 24-31. https://doi.org/10.1177/0098628316677643

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *The American Journal of Psychology*, *9*(4), 507–533. https://doi.org/10.2307/1412188

Tucha, L., Fuermaier, A. B., Koerts, J., Buggenthin, R., Aschenbrenner, S., Weisbrod, M., Thome, J., Lange, K. W., & Tucha, O. (2017). Sustained attention in adult ADHD: Time-on-task effects of various measures of attention. *Journal of Neural Transmission*, *124*(Suppl 1), 39–53. https://doi.org/10.1007/s00702-015-1426-0

Uziel, L. (2007). Individual differences in the social facilitation effect: A review and meta-analysis. *Journal of Research in Personality, 41*, 579-601. https://doi.org/10.1016/j.jrp.2006.06.008

Vodanovich, S. J., & Watt, J. D. (2016). Self-report measures of boredom: An updated review of the literature. *The Journal of Psychology*, *150*(2), 196–228. https://doi.org/10.1080/00223980.2015.1074531

Velasco, F., & Jorda, R. (2020). Portrait of boredom among athletes and its implications in sports management: A Multi-method approach. *Frontiers in Psychology*, *11*, 831. https://doi.org/10.3389/fpsyg.2020.00831

Warm, J.S. (1977). Psychological processes in sustained attention. In: Mackie, R.R. (Ed.) Vigilance (Vol. 3). NATO Conference Series. Springer. https://doi.org/10.1007/978-1-4684-2529-1_30

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors, 50*(3), 433-441. https://doi.org/10.1518/001872008X312152

Westinghouse Electric Corporation. 1981. *Product safety label handbook.* Westinghouse Printing Division.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Wogalter, M. S., Allison, S. T., & Mckenna, N. A. (1989). Effects of cost and social influence on warning compliance. *Human Factors*, *31*(2), 133-140. https://doi.org/10.1177/001872088903100202

Wogalter, M. S., Brelsford, J. W., Desaulniers, D.R. and Laughery, K. R. (1991). Consumer product warnings: The role of hazard perception. *Journal of Safety Research, 22,* 71-82. https://doi.org/10.1016/0022-4375(91)90015-N

Wogalter, M. S., Brems , D. J. and Martin, E. G. (1993). Risk perception of common consumer products: Judgments of accident frequency and precautionary intent. *Journal of Safety Research, 24,* 97-106. https://doi.org/10.1016/0022-4375(93)90004-7

Wogalter, M. S., Godfrey, S. S., Fontenelle, G. A., Desaulniers, D. R., Rothstein, P. R., & Laughery, K. R. (1987). Effectiveness of warnings. *Human Factors*, *29*(5), 599-612. https://doi.org/10.1177/001872088702900509

Wogalter, M. S., Kalsher, M. J ., Frederick, L. J., Magurno, A. B. and Brewster, B. M. (1998a). Hazard level perceptions of warning components and configurations. International *Journal of Cognitive Ergonomics, 2*, 123-143.

Wogalter, M. S., Magumo, A. B., Dietrich, D. and Scott, K. (1999e). Enhancing information acquisition for over-the-counter medications by making better use of container surface space. *Experimental Aging Research*, *25*(1), 27–48. https://doi.org/10.1080/036107399244129

Wogalter, M. S. (2006b). Communication-human information processing (C-HIP) model. In M.S. Wogalter (Ed.), *Handbook of Warnings* (pp. 51-61). Lawrence Erlbaum Associates Publishers.

Yamashita, A., Rothlein, D., Kucyi, A., Valera, E.M., Germine, L., Wilmer, J., DeGuitis, J., Esterman, M. (2021). Variable rather than extreme slow reaction times distinguish brain states during sustained attention. *Scientific Reports, 11*(14883). https://doi.org/10.1038/s41598-021-94161-0

Zajonc, R. B. (1965). Social facilitation. *Science, 149*(3681), 269-274. https://doi.org/10.1126/science.149.3681.269

Zajonc, R. B., & Sales, S. M. (1966). Social facilitation of dominant and subordinate responses. *Journal of Experimental Social Psychology*, *2*(2), 160–168. https://doi.org/10.1016/0022-1031(66)90077-1