WORRY AND WORKING MEMORY: A BEHAVIORAL & ERP INVESTIGATION ACROSS THE MENSTRUAL CYCLE

By

Lilianne Marie Gloe

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology—Master of Arts

2019

ABSTRACT

WORRY AND WORKING MEMORY: A BEHAVIORAL & ERP INVESTIGATION ACROSS THE MENSTRUAL CYCLE

By

Lilianne Marie Gloe

The current project examined the relationship between worry, a component of anxiety characterized by negative, future-oriented thought activity, and working memory in women. Further elucidation of this relationship occurred through use of multiple working memory measures; examining the P300, an event-related potential measured using electroencephalogram (EEG) thought to index resources available for cognitive processing; and considering the role of ovarian hormones. It was hypothesized that worry would be associated with poorer working memory function and reduced amplitude of the P300 at higher levels of task difficulty and estradiol. Participants were 65 naturally-cycling women who attended four visits across their menstrual cycles. On each visit day, data collection included a measure of daily worry (Penn State Worry Questionnaire), a saliva sample used to assay for estradiol, and completion of three working memory tasks (*N*-back task with EEG recording, Operation Span task and Reading Span task). Five multilevel models were constructed to examine the impact of within-subject fluctuation of Penn State Worry Questionnaire scores and estradiol values on *N*-back task accuracy, *N*-back task reaction time, *N*-back task P300 amplitude, Operation Span score, and Reading Span score. Task parameters of the *N*-back task (i.e. load and trial type) were included in the three models of the *N*-back task as indicators of task difficulty. Results indicated that within-subject fluctuations were not significantly related to working memory performance or the P300 amplitude. Potential reasons for null findings and future directions are explored.

ACKNOWLEDGEMENTS

I would like to thank Deborah Kashy for her valuable consultation with regards to the multilevel modeling structure utilized for the *N*-back task, as well as the Clinical Psychophysiology Lab undergraduate research assistants, lab managers and graduate students for their contributions to data collection for this project. Additionally, I would like to thank my Master's committee members, Dr. Katharine Thakkar, Dr. Susan Ravizza and Dr. Jason Moser, for their feedback on this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Sex disparities in anxiety have long been recognized. Women suffer from higher rates and longer course of anxiety than men (Kessler, Petukhova, Sampson, Zaslavsky, & Wittchen, 2012; McLean, Asnaani, Litz, & Hofmann, 2011). Despite the disproportionate number of women suffering from the effects of anxiety, studies have largely failed to examine the precise nature and impact of anxiety in women. Thus, the current study focused on the relation between anxiety and one important aspect of functioning previously linked to anxiety– cognition – in women.

Models of anxiety and cognition are currently stated in general terms and are intended to apply to all people, regardless of sex. One prominent model of the relationship between anxiety and cognitive impairment is Attentional Control Theory (ACT; Eysenck & Calvo, 1992; Eysenck & Derakshan, 2011; Eysenck et al., 2007). The theory asserts that anxiety, via worrisome thoughts that load the cognitive system, depletes available working memory resources. To overcome, or compensate for, this dearth of working memory resources, ACT proposes that anxious individuals recruit auxiliary resources so that they can complete the task at hand (Eysenck & Calvo, 1992; Eysenck et al., 2007). As a result, they demonstrate inefficient performance, expending greater resources to complete the task at a comparable level of accuracy as non-anxious individuals (Eysenck et al., 2007).

While ACT has garnered empirical support (e.g. Eysenck et al., 2007), the literature is limited and heterogeneous. In particular, whether and how worry – a component of anxiety comprised of negative, future-oriented verbal thought activity (Borkovec, Ray, & Stober, 1998) -- is related to deficits in working memory is unclear. The applicability of this theory to women has also been assumed, but has not been explicitly examined. In addition to being central to

ACT, understanding how worry relates to working memory is particularly critical for women, given that women have been shown to experience higher levels of worry than men (Nitschke, Heller, Imig, McDonald, & Miller, 2001). Therefore, the current study aimed to better characterize the relationship between worry and working memory in women. To further elucidate this important relationship, this study 1) utilized several working memory tasks, 2) included behavioral and neurophysiological measures, and 3) considered the role of ovarian hormones in women across the menstrual cycle.

A Primer on Working Memory

Working memory (WM) is a multi-component system that reflects one's ability to simultaneously store and manipulate information during tasks (Baddeley, 1996; Baddeley, 1986). WM is thought to be crucial for "goal-directed behavior" in that it allows goal-relevant information to be readily available despite competing distractors (Kane, Conway, Hambrick, & Engle, 2007). Given that WM is a multi-faceted concept, many tasks have been created to measure it. For the purposes of this brief review, I will focus mostly on literature that uses two common types of tasks: dynamic span and complex span tasks.

Dynamic span tasks.

Dynamic span tasks involve continuous attention to a series of items presented one after another, and the updating of WM to reflect relevant target items during presentation (Moran, 2016). The *N*-back task is one of the most common ways in which to measure dynamic span. It requires participants to respond when an item is shown that had been presented *N* trials previously, with *N* defined for each block of trials. *N*-back tasks exist in several modalities, including a verbal version in which participants respond when a particular letter is shown visually that matched the one shown N trials earlier, and a spatial form, in which participants

respond when the letter is shown *in the same location* as was shown N trials earlier. In addition to manipulating task load by varying the number of trials back the target letter occurred in the sequence (i.e. *N*), variations of this task have also modulated task difficulty through the inclusion of lure trials within the sequences. Lures are letters that match a recently shown, but non-target letter. Lure trials are thought to produce more interference than other non-targets, as inhibiting responses to lures requires more cognitive control because of the familiarity/saliency of stimuli (Gray, Chabris, & Braver, 2003).

Complex span tasks.

In complex span tasks, participants are presented with a series of letters interleaved with a demanding secondary task (Moran, 2016). Two popular complex tasks are the Operation-Span (O-Span; Turner & Engle, 1989; Unsworth, Heitz, Schrock, & Engle, 2005) and Reading-Span tasks (R-Span; Daneman & Carpenter, 1980), in which participants are presented with letters interleaved with mathematical operation problems (O-Span) or sentence comprehension exercises (R-Span) for a sequence of trials. At the conclusion of the trial sequence, they are asked to recall the presented letters in perfect order. The sum of the number of letters contained within perfectly recalled sequences are used to produce a WM capacity score.

A Brief Review of the Relationship Between WM and Worry

Although a great deal of work has investigated the relationship between anxiety and WM processes, the nature of the more specific relationship between worry and WM processes is still unclear. A recent meta-analysis examined the worry and WM relationship for tasks in the spatial (k=3, N=258) and phonological domain (k=7, N=647) and found moderate and small effect sizes, respectively (Moran, 2016). However, in examining the worry and WM literature more closely, the few studies that have examined this relationship are very diverse in methodology.

They utilize a variety of WM tasks, worry measures, sample compositions (e.g. clinical v. healthy participants, age), and study designs (e.g. worry induction vs. trait worry). Because variability in the methods and sample composition is high, especially in proportion to the number of studies of worry and WM overall, it is important to understand the evidence for this relationship in the context of individual studies.

In healthy populations, studies utilizing dynamic span tasks have found relationships between trait worry and lower accuracy on 2-back and 3-back versions of the *N*-back task (Bredemeier & Berenbaum, 2013; Moran, 2016). Additionally, one study of participants diagnosed with Generalized Anxiety Disorder (GAD) – a disorder characterized by excessive worry (American Psychiatric Association, 2013) – showed that GAD patients had increased reaction times on 2-back and 3-back correct target trials of a verbal N-back task (Stefanopoulou, Hirsch, Hayes, Adlam, & Coker, 2014). Thus, although there appears to be a relationship between worry and dynamic span performance, the way in which worry impacts behavioral performance is unclear.

A more obscure relationship emerges in studies utilizing complex span tasks. In the same study that found a relationship between lower *N*-back accuracy and trait worry, Bredemeier & Berenbaum (2013) found no relationship between worry and O-Span score. In contrast, Trezise & Reeve (2014) found that state testing worry was associated with lower accuracy on a modified O-Span task that used arithmetic problems as target stimuli in adolescent students. Ganley & Vasilyeva (2014) further showed that testing worries were related to poor performance on two different span tasks. However, in a second study, which included a larger sample, a more balanced sex distribution, and was conducted at a different university, they only observed a relationship between worry and poor visuospatial WM (Ganley & Vasilyeva, 2014). Although it

is unclear why the relationship between worry and complex span task performance differed across studies, it could be related to the sample characteristics (i.e. adults v. adolescents; female dominated v. equal sex distribution), the difference in worry measures (i.e. trait v. state), and/or the use of different tasks and stimuli.

The Current Study

Despite previous literature demonstrating that the relationship between worry and WM appears to be present across many contexts, the heterogeneity in results indicates the need for further work to understand the precise nature of this relationship. This need was addressed in the current study by considering multiple measures of WM, a neurophysiological index of WM and the role of ovarian hormones in women across the menstrual cycle.

Multiple behavioral measures of WM.

First, I examine how worry is related to performance on three measures of WM. Specifically, I have included a verbal *N*-back task and two complex span tasks in the analysis to more fully capture WM as multi-modal construct in its relationship with worry. In exploring dynamic and complex span tasks together, recent investigations have found that, while both tasks seem to assess general WM, unique task-specific factors provide additional information that cannot be yielded through the use of a single WM task alone. For instance, results from a meta-analysis found weak correlations between *N*-back and complex span performance (Redick & Lindsey, 2013). However, in response to the findings of Redick & Lindsey (2013), a study using latent-variable analysis found strong correlations between task types after controlling for measurement error and task content (e.g. letters vs. numbers; Schmiedek, Lövdén, & Lindenberger, 2014). At the neural level, Minamoto et al. (2017) found that the lateral prefrontal cortex (lPFC) appears to be implicated in task-related functional connectivity across both *N*-back

and complex span tasks. However, the regions that are recruited in tandem with the lPFC differ by task and *N*-back content (Minamoto et al., 2017). Together, the extant literature suggests that dynamic and complex span tasks index a general aspect of WM while simultaneously assessing elements of WM that are task-specific.

Because task-specific variance is important to consider in studies of WM, I have utilized both dynamic and complex span tasks in the current analysis to allow for assessment of task-general and task-specific effects of worry on WM. Task-general WM is more global and does not depend on the content of the to-be maintained/manipulated items, while task-specific WM is specific to the content of such items and can be divided into verbal/phonological and visuospatial content categories. Recent work has provided evidence for a relationship between worry and reduced task-general WM (Moran, 2016). However, the verbal nature of worry suggests it could also relate to poorer processing of verbal/phonological items, in particular, above and beyond general WM deficits (Borkovec et al., 1998; Moran, 2016). Importantly, the tasks used in the current study all utilize verbal stimuli to be manipulated. It should be noted that R-Span task may introduce additional interference in verbal-related processes because the secondary task involves verbal material, as opposed to the numerical operations utilized by the O-Span task.

Neurophysiology.

Second, although examining behavioral performance on several WM tasks will contribute to a fuller understanding of how worry may relate to poor WM function, additional information about the nature of this relationship can be gleaned from neurophysiological investigation. To date, no work has directly examined the relationship between worry and neural correlates of WM.

Using electroencephalogram (EEG), a direct measure of online electrical brain activity, neural responses during WM tasks can be examined with high temporal precision. These neural responses, called Event-Related Potentials (ERPs), consist of deflections of voltage that occur in response to a variety of external (e.g. presentation of a digit) and internal (e.g. commission of an error) stimuli and reflect specific neural or psychological processes (Kappenman & Luck, 2011). Although the relationship between worry and WM deficits has not been examined using neurophysiological methods, there is precedence for recording ERPs during the *N*-back task. Past work has indicated that the amplitude of the P300 (or P3b) ERP component is modulated by level of load in the *N*-Back task (Bailey, Mlynarczyk, & West, 2016; McEvoy, Smith, & Gevins, 1998; Watter, Geffen, & Geffen, 2001; West, Bowry, & Krompinger, 2006). The P300 is a positive voltage change maximal at parietal sites that peaks approximately 300 – 800ms after a task-relevant, salient, or novel stimulus is presented. The P300 can be thought of as an index of attentional resource allocation, such that it is increased when more attention is devoted to the relevant external stimuli (i.e. target trials; Polich, 2012). This interpretation can be applied to the existing literature demonstrating that the P300 is reduced in amplitude as memory load increases in the N-Back (Bailey et al., 2016; McEvoy et al., 1998; Watter et al., 2001; West et al., 2006). At higher loads, there is a great reliance on WM processes, because more stimuli must be held in WM and updated as new stimuli are presented. Thus, the reduced P300 amplitude at higher loads reflects fewer available resources for processing the external stimuli, because such resources are being devoted to internal WM processes. Because worry puts further strain on WM resources, the P300 should be even smaller at higher loads in worriers.

Sex and ovarian hormones.

Finally, the current study also considers the relationship between worry and WM in women as a function of hormonal status across the menstrual cycle. It is possible that the heterogeneity of previous worry and WM findings may also be due, in part, to ignoring sex differences and/or fluctuations of ovarian hormones across the menstrual cycle.

It is critical to consider the role of ovarian hormones in studies of cognition and anxiety, given their role in both (Gasbarri et al., 2008; Jacobs & D'Esposito, 2011; Li & Graham, 2017; Maeng & Milad, 2015; Man, MacMillan, Scott, & Young, 1999; Montoya & Bos, 2017). Animal work by Shansky and colleagues have found that female rat WM performance suffers under acute stress at high levels of estradiol. The worst performance is found in (1) naturally cycling rats during the proestrus phase – where estradiol is high -- compared to rats in phases characterized by lower estrogen levels and (2) in overiectomized rats given a long-term estrogen replacement (Shansky & Lipps, 2013; Shansky et al., 2004; Shansky, Bender, & Arnsten, 2009; Shansky, Rubinow, Brennan, & Arnsten, 2006). Hypothesized mechanisms by which high estradiol may negatively impact WM performance under stress include the intensification of the effects of glutocorticoid release, increases in the availability of dopamine, and disruption of the balance between dopaminergic and noradrenergic receptor activation (Shansky & Lipps, 2013).

Some human work has also reported deleterious effects of stress on WM in women (Schoofs, Pabst, Brand, & Wolf, 2013), but findings have been inconsistent and have not examined the role of ovarian hormones (Shields, Sazma, & Yonelinas, 2016). Moreover, despite the examination of stress and WM in females, to my knowledge, no studies have examined the role of ovarian hormones in the relationship between worry and WM.

Hypotheses.

The primary hypothesis of the current study was that worry would be associated with poorer working memory function at higher levels of difficulty and estradiol, because it is under these conditions that worry's depleting effects are proposed to be greatest (Eysenck et al., 2007; Shansky & Lipps, 2013). It was expected that worry would be related to deficits in WM performance (either in longer RT or reduced accuracy) and a reduced P300 amplitude when the task was most difficult (i.e. at higher loads and more difficult trial types of the N-back), but only when estradiol was high. Because it was unclear given past literature if worry's impact on the N-back task is task-general, two hypotheses for how estradiol may be implicated in the relationship between worry and complex span task performance were considered. If worry relates to task-general WM, then worry would be related to poorer performance on both span tasks when estradiol is high. If worry has a particularly strong effect on phonological/verbal WM, then worry would be related to poorer performance on the R-Span only.

METHODS

Participants

Participants were 67 female volunteers 18 to 25 years old recruited for the MSU Clinical Psychophysiology Lab Brain Cycle Study. Participants were recruited from East Lansing, Lansing and surrounding areas in Michigan via commercial mailing lists, paid and public service advertisements in local media, flyers, and online advertisements via Craigslist. Demographic information about the sample is provided in Table 1. Overall, the sample was predominately white, heterosexual and had a gender identity of female. However, notably, the sample also consisted of several participants who belong to racial and sexual orientation groups that are traditionally under-represented/reported in the literature.

Because the Brain Cycle Study aims to examine cognitive impacts of worry in women across the menstrual cycle, many inclusion/exclusion criteria are in place to ensure other exogenous and endogenous factors are not acting upon endocrine system functioning. Thus, to be eligible for the study, participants had to be naturally menstruating (i.e. every 22-35 days) and not taking hormonal contraceptives, psychotropic medications and steroid medications during the past eight weeks before study participation. Women must have had no history of genetic or medical conditions known to have an impact upon the endocrine system. Additionally, those who had epilepsy; have hearing, visual, or other physical or mental impairments that could interfere with data quality; or had experienced head trauma that resulted in a loss of consciousness for over five minutes were also excluded from the study because of potential effects on neurophysiological data collection. Eligibility was confirmed during each visit of the study to ensure criteria was consistently met.

Procedure

Overview.

The Brain Cycle Study consists of daily questionnaires and saliva sample collection, one intake visit, four EEG visits and a final visit for administration of a structured clinic diagnostic interview. An overview of data collection is provided in Figure 1.

Volunteers interested in participating in the study were screened over the phone for eligibility using the aforementioned criteria. Menstrual cycle history was also collected during the phone screen so that the study staff could schedule participants for EEG visits that corresponded to phases of their menstrual cycle, enabling the collection of data across the entire menstrual cycle for each participant. The timing of the first EEG visit was randomly selected based on their current menstrual cycle phase (i.e. early follicular phase, late follicular phase, ovulation phase, or mid-luteal phase) to ensure that a similar number of participants started in each of the four phases.

Eligible participants came to the Clinical Psychophysiology Lab (CPL) in the MSU Psychology Building for an intake visit that consisted of confirmation of eligibility criteria, orientation to study procedures, the provision of important information regarding daily data collection procedures (e.g. the website address for daily online questionnaire assessments), and carrying out written consent procedures. Participants were compensated $280 for full participation, with prorated compensation provided for partial data collection.

Daily questionnaire procedure.

Participants were asked to fill out a series of questionnaires between 5:00PM and 10:00PM each day via Qualtrics, an online assessment portal. Paper copies of the questionnaires were provided to each participant in case they are unable to access the Internet during this

timeframe. The daily questionnaires consisted of the Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990) and other assessments of anxiety, depression and eating pathology that were not included for the current analysis. The PSWQ asked about worry during the day that the questionnaire was completed.

Daily saliva sample collection.

After being provided with collection tubes at the intake visit, participants were taught to use the passive drool method to provide their daily samples. Participants were instructed to provide their samples within thirty minutes of waking and were asked not to eat, drink, brush their teeth, or smoke before samples are taken. Samples were provided for 35 consecutive days in order to capture estradiol levels across the menstrual cycle. Participants were instructed to store their samples in their home freezer immediately after daily collection and were provided with materials to ensure their samples did not thaw when transported to the lab. Samples were logged and stored in a lab -80 °F freezer until they were shipped to Salimetrics, LLC (State College, PA). Saliva was analyzed using enzyme immunoassay kits for assaying estradiol as specified by (Klump et al., 2016). Of samples collected on EEG visit days, a 99% retention rate was achieved for estradiol assay.

EEG visit procedures.

Each participant reported to the CPL in the MSU Psychology building for each EEG visit, estimated to last two to three hours in duration. After confirming eligibility, study staff completed set-up for EEG recording. The participant then completed a battery of cognitive assessments on the computer, which consisted of a Flanker task (not included in the current analysis), followed by verbal *N*-back task and automated O-Span and R-Span tasks. The O-Span and R-Span tasks were counterbalanced for order of administration. EEG recoding was only

conducted during the Flanker and *N*-back task. At the end of the visit, the participant completed an online questionnaire assessment. This assessment contained a series of questionnaires assessing for a wide variety of attitudes, beliefs and symptomology that were not used in the current analysis.

N-Back task.

The verbal *N*- back task, constructed by Jacobs & D'Esposito (2011), involves the presentation of a continuous stream of letters one at a time in a time-locked fashion. Each letter is displayed for 1000ms, with an ITI of 1100ms. Participants were asked to respond within 2000ms of the stimulus appearing using the two mouse buttons, with each mouse button corresponding to targets and non-targets. Memory load was manipulated such that the target for each block of trials is identified as the letter shown *n* trials previously in the block. Targets were *0*-back, *2*-back or *3*-back letters. During *0*-back trial blocks, participants were asked to respond to the target letter "X" whenever it appears. In *2*-back or *3*-back trial blocks, the target letter was the letter that appeared two or three trials, respectively, before the current trial. For example, in the sequence A-T-R-T-C, the second T in the sequence is the target in a 2-back block. The task consisted of 16 blocks with a total of 320 trials (*0*-back = 160 trials, *2*-back = 80 trials, *3*-back = 80 trials). In addition to standard target and non-target trials, a small subset of non-target trials presented during *2*-back and *3*-back blocks consisted of lure trials. Lures are letters that match a recently shown, but non-target, letter. As an example, consider the sequence K-F-E-D-K shown during a *3*-back trial block. The second "K" presented is a lure trial, because K was originally presented four letters back, not three letters back. Participants were counterbalanced to one of four versions of the N-back using a Latin Square design.

N-back data was preprocessed in Matlab. Practice trial were removed from the data set. Trials were included in calculations of reaction time and accuracy if the reaction time was greater than 200ms. Reaction time was only calculated for trials on which correct responses were given.

O-Span and R-Span tasks.

The automated O-Span task (Turner & Engle, 1989; Unsworth et al., 2005) consisted of a series of mathematic operation problems, whereas the automated R-Span task (Daneman & Carpenter, 1980) involved a series of comprehension exercises. In between each math problem or sentence comprehension, a letter was presented. At the end of the series of trials, participants were required to recall the letters in their presented order. The number of letters presented ranged from 3 to 7 for each series of trials. For each task, a final score was calculated as the total number of letters recalled in series for which recall was perfect for all letters in the series. The maximum final score was 75. Scores were calculated in E-Prime.

Neurophysiological data recording and preprocessing.

Continuous electroencephalographic (EEG) activity was recorded during the N-back task using the ActiveTwo Biosemi system (BioSemi, The Netherlands) from 64 Ag-AgCl electrodes fitted in a stretch-lycra cap. The location of the cap electrode ports is based on the 10-20 system. The "10-20 system" refers to the standardized method of placing each of the scalp electrodes – each electrode is spaced apart from adjacent electrodes at a distance of either 10% or 20% of the total front-back to right-left distance of the skull. Measurements were taken to ensure proper cap fit, with cap size determined by the distance from the nasion (the distinctly depressed area between the eyes) and the inion (the lowest point of the skull on the back of the skull identified by a prominent bump). Centering of the cap was achieved by measuring the distance between the

ears around the top of the head, with the tip of each ear being used as a measurement endpoint. A chin strap was used to hold the cap in place in a tight, but comfortable fashion. Electrodes were plugged into each of the labeled ports, with labels consisting of combinations for letters and digits (e.g. Pz, C2, T7). The first letter of the label corresponds to areas of the cerebral cortex (i.e. F = frontal, T= temporal, C= central, P = parietal, and O = occipital lobes). The second part of the label can either be a letter or number and indicates location on the scalp in relation to midline sites. The letter "z" indicates a location along the midline of the scalp, while odd numbers indicate left hemisphere sites and even numbers indicate right hemisphere sites.

Sensors were also placed on the left and right outer canthi (the outer corners of the eyes where the upper and lower lids meet) and below the left eye (approximately 1cm from the pupil). Activity recorded from the FP1 site and the three external eye sensors were used to measure electrooculogram (EOG) activity resulting from blinks and eye-movements. Two sensors were also placed on the left and right mastoids – bone protrusions behind the ears – to use during offline analyses as references. The Common Mode Sense (CMS) active electrode and the Driven Right Leg (DRL) passive electrode formed the ground during data acquisition. In addition to acting as a reference, the CMS-DRL loop ensures that the average voltage of the participant stays within a reasonable range, thereby limiting current that could potentially return to the participant. All signals were digitized at 1,024Hz, which represents 1,024 samples of data taken per second and provides millisecond precision.

Offline Analyses was conducted with BrainVision Analyzer 2 (Brain Products, Gilching, Germany). Recordings were visually examined by L. Gloe to determine if problems during EEG recording resulted in significant artifactual noise at any recording channel. If a small number (less than or equal to five) channels were found to contain obvious noise, these channels were

removed and interpolated based on activity in the channels closest to the removed channel on the scalp. If greater than five channels needed to be interpolated for adequate data to be obtained, the participant's EEG recording was not included in analyses and was considered missing. Recordings were band-pass filtered with cutoffs of 0.1 Hz and 30Hz (12dB/oct roll off) and re-referenced to the numeric mean of the mastoids. Ocular correction was then conducted using a common regression method (Gratton, Coles, & Donchin, 1983). This method accounts for eye movement and blinks, which are typically at their greatest magnitude at frontal sites of the scalp that are near the eyes. Additionally, the method includes calculation of a propagation factor that estimates the differential impact of these movements at sites across the scalp. The recordings were then segmented based on cognitive load (0-back, 2-back, 3-back) and stimulus type (target, nontarget, lure) on correct trials only – error trials were removed. Segments were made relative to stimulus presentation, such that segments begin 200ms prior to each letter stimulus and terminate 1,000ms post-stimulus onset. For each set of segments, artifact correction was carried out using an algorithm such that trials were rejected if they contained the any of the following: activity characterized by a voltage step greater than 50 microvolts/ms compared to both the preceding and following trials, respectively; a 200ms time window with a difference in voltage of 300 microvolts; a 100ms time window in which a difference in voltage was less than 0.5 microvolts; or an amplitude more extreme than +/-200 microvolts. Within in each segment, activity in individual channels was averaged across trials, resulting in a single average for each channel of each segment for each participant. Baseline correction subtracted the average activity 200ms prior to stimulus onset from each data point after stimulus onset for each trial type. These averages for each participant were then averaged across participants to create a "grand-average".

The P300 was then scored in the 300-500ms time window at Pz, where it reached maximal amplitude.

Analysis Procedure

Data were cleaned and prepared for modeling in R (Version 3.4.2). PSWQ score and estradiol were person-mean centered, such that the mean for each measure was calculated across visits for a subject and subtracted from that subject's measure/score on each day. Thus, significant effects involving these measures in models can then be interpreted in terms of fluctuations *within* a person instead of in terms of differences between people. Additionally, accuracy was examined across each load-by-trial-type interaction level and was labeled as missing for a level if it fell below 30%, as this is believed to be an indicator of lack of effort or misunderstanding task directions.

Multilevel models were utilized in order to examine the impact of worry and estradiol on *N*-back task accuracy, *N*-back task reaction time, the P300 elicited during the *N*-back, O-Span Score, and R-Span score over four EEG visits across the menstrual cycle. All modeling was executed in SAS software (Version 9.4; SAS System for Windows) to accommodate the necessary model structure. For all multilevel models, a random intercept was included to account for nonindependence due to repeated measures of behavioral measures, the PSWQ, estradiol, and P300 amplitude from each subject. Intra-class correlations (ICCs) were calculated for each model to provide a measure of the extent to which behavioral measures and the P300 amplitude from the same subject were correlated. Additionally, the proportion of variance explained by each model (i.e. marginal R^2) was calculated using residual variance from the base/unconditional model, which contained no fixed effects and contained only the intercept, and comparing it to that of the full model for each dependent variable. The following formula was utilized: R^2 =

$\frac{\sigma_b^2 - \sigma_f^2}{\sigma_b^2}$, where σ_b^2 represents the residual variance of the base model and σ_f^2 represents the residual variance of the full model.

For the fixed effects of all multilevel models constructed, the effect of EEG visit number was included as a covariate in order to account for any practice effects that may have occurred as a result of completing the *N*-back and span tasks multiple times. The significance of each variable and their interactions was assessed using a Type III test for fixed effects, which test degree of unique variance explained by a categorical predictor over and above all other variables in the model.

***N*-back performance and P300 analyses.**

For the *N*-back task, three models were constructed that varied by their dependent variable, such that there was a model for accuracy, reaction time, and the P300 at Pz from 300 – 500 ms. The site and time window for the P300 were selected based on common sites examined in the literature and based on visual examination of grand-mean averages of the data.

The random effects for the models of the *N*-back task were constructed to account for heterogeneity in variance across and covariances between measures at each of the eight levels of the load-by-trial-type interaction (i.e. 0-back non-target, 0-back target, 2-back non-target, 2-back target, 2-back lure, 3-back non-target, 3-back target, 3-back lure) across each EEG visit within each participant. Load-by-trial type levels occurring within the same visit and subject would be expected to be more similar than those occurring in separate visits or in different subjects. An unstructured covariance matrix was utilized, given evidence of considerable differences in variances and covariances within and between load-by-trial-type interaction levels, respectively. In order to calculate ICCs and the proportions of variance explained by the full model (i.e. R^2),

an estimate for residual variance was obtained by averaging the estimates of variance for each of the eight trial types.

For each *N*-back model, a 4-way interaction was tested between load, trial type, estradiol and PSWQ scores in order to understand the impact of estradiol and PSWQ scores on the *N*-back task measures at different task difficult. Examining this interaction allowed me to test the hypothesis that higher levels of worry measured via PSWQ scores would be related to increased RT, decreased accuracy, and a smaller P300 at higher levels of load (i.e. 2-back and 3-back loads) and on more difficult trial types (i.e. target and lure trials; see Gray, Chabris, & Braver, 2003) when estradiol was high.

All lower-order terms were included in each "full model" in addition to a main effect of EEG visit number. Estimates for load, trial type and their interaction were derived from the least squares means from the full model, which represents the partial means for each level of each variable/interaction while holding all other variables in the model at their mean. Estimates for estradiol, PSWQ scores, and their interaction were evaluated by running the same full model using effects-coded predictors for trial type and load. Any significant interactions between continuous and categorical variables were broken down by simple slopes analyses in which separate intercepts and slopes for the continuous variable were computed at each level of the categorical variable(s) involved. Any interaction involving both estradiol and PSWQ score was followed-up by using the procedures described by Aiken, West, & Reno (1991). Simple slopes analyses were conducted to examine the effects of one continuous variable at high (+1SD) and low (-1SD) values of the other continuous variable.

R-Span and O-Span score analyses.

For the models of span scores, the interaction between estradiol and PSWQ was tested along with its lower order terms and a main effect of EEG visit number. Models were created with person-mean centered PSWQ scores and estradiol as covariates.

If worry impacts task-general WM, PSWQ was expected to be related to reduced O-Span and R-Span scores when estradiol is high. On the other hand, if worry is specifically implicated in verbal/phonological-specific WM, PSWQ score would be related to reduced R-Span scores at higher levels of estradiol. There would be no relationship between worry and O-Span scores in this scenario.

RESULTS

Data Retention

Figure 2 depicts a flowchart of the data loss across the various measures. The number of subjects with at least three sessions worth of data for analyses was 61 for *N*-back accuracy and reaction time, 59 for EEG analyses, and 61 for O-Span and R-Span. For the sake of power, subjects with sessions missing daily estradiol and PSWQ Score on some (but not all) visit days were retained for analyses, although their data on the missed day was not analyzed.

Descriptive Statistics for Predictors and Dependent Variables

Descriptive statistics for PSWQ scores, estradiol, and R-Span and O-Span scores are depicted in Table 2. Descriptive statistics for *N-back* measures are provided are provided in Table 3. ERP waveforms for each trial type at Pz are presented in Figure 3.

N-Back Task

Accuracy.

Results of the Type III tests for the fixed effects of the full model predicting accuracy are presented in Table 4. First, the influence of trial type and load on *N*-back accuracy was observed. By examining the effect of task-variables on performance irrespective of PSWQ scores and estradiol, a clearer understanding of how the *N*-back task functions over time can be achieved. Notably, variances at each trial type-by-load level differ greatly from one another, as do the correlation of accuracy between levels over time (Table B1). This suggests that accuracy at some levels of the task changes more across time than at other levels *and* that accuracy at levels of the task are uniquely related to each over time. Model estimates for main effects and least square means for trial-by-load interaction levels are displayed in Tables B2 and B3. Significant differences in accuracy were found between all trial type (p <0.001) except between 2-back non-

targets and 3-back non-targets ($t(223) = 1.99$, $p = 1.00$), 2-back target and 2-back lures ($t(220) = 2.42$, $p = 0.453$), 3-back targets and 3-back lures ($t(212) = 0.02$, $p = 1.00$). Accuracy significantly reduced as load increased for targets and lures, while accuracy on non-target trials did not differ between 2-back and 3-back loads. As expected, increased load lead to worse performance on more difficult trial types (i.e. targets and lures), but not on the easier non-target trials.

In examining interactions involving PSWQ score and estradiol, marginal interactions PSWQ x estradiol and PSWQ x estradiol x load emerged. In breaking down the latter three-way interaction, it was discovered that this interaction was driven by a significant simple slope for PSWQ Scores at low levels of estradiol on 3-back trials ($B = -0.2159$, $SE = 0.096$, $t(209) = 0.025$). As depicted in Figure 4, higher PSWQ scores were related to lower 3-back accuracy in the presence of low estradiol levels. Results support the hypothesized relationship between higher PSWQ scores and lower *N*-back performance on harder trial types, but are in direct contrast to the hypothesis that such a relationship should emerge in the presence of high estradiol levels.

Reaction time.

Type III test of significance for the full model for reaction time are presented in Table 5. First, the influence of trial type and load on N-back reaction time was observed. Similar to the model for accuracy, residual variances at each trial type-by-load level differ greatly from one another, as do the covariances between residuals of levels over time (Table B4). This suggests that reaction time at some levels of the task changes more across time than at other levels *and* that reaction time at levels of the task are uniquely related to each overtime. Model main effects and least square means for trial-type-by-load interaction levels are displayed in Tables B5 and B6. As shown in Table 5, significant effects of trial type, load and their interaction emerged from

the full model. Significant differences were found between all trial types (p <0.05) except between 3-back targets and 2-back lures ($t(225) = -1.03$, $p = 1.00$), 2-back target and 3-back lures ($t(225) = -1.84$, $p = 1.00$), 3-back targets and 3-back lures ($t(225) = 2.16$, $p = 0.8985$). Together, results suggest that participants respond more slowly as load increases for non-targets and targets and respond most quickly on non-target trials compared to targets and lures. This pattern of responding confirms my expectations in terms of more slow responses for more difficult trial types, although the lack of significant differences between 2-back and 3-back loads for targets and lures was surprising.

Importantly, the effect of estradiol and PSWQ score was also examined in the model. Contrary to my hypothesis, however, PSWQ scores and estradiol levels were not significantly related to *N*-back reaction time.

P300 at Pz 300-500 ms.

Results of Type III test of significance for the full model for the P300 in the time window 300-500ms at site Pz are presented in Table 6. First, the influence of trial type and load on N-back reaction time was observed. Similar to reaction time and accuracy models, variance and correlation estimates across time were heterogeneous by load by trial type interactions (Table B7). Again, this suggests that P300 amplitude at some levels of the task changes more across time than at other levels *and* that P300 amplitude at levels of the task are uniquely related to each over time. Model estimates for main effects and least square means for trial type-by-load interaction levels are displayed in Tables B8 and B9. Significant differences between least squares means for the levels of this interaction are presented in Table 7. As expected, the P300 amplitude was significantly reduced on 2-back and 3-back loads compared to 0-back load for targets and on 3-back load compared to 2-back load for non-targets. However, in contrast to

expectations, P300 amplitude did not significantly differ between 2-back and 3-back targets nor between 2-back and 3-back loads compared to 0-back loads for non-targets. Additionally, in contrast to expectations, load did not influence P300 amplitude on lure trials.

Importantly, the effect of estradiol and PSWQ score was examined. Contrary to hypotheses, PSWQ scores and estradiol levels were not significantly related to P300 amplitude.

Span Tasks

R-Span task.

Results of the full model for R-Span score are presented in Table 8 and estimates for the intercept, estradiol, PSWQ score and their interaction are depicted in Table B10. No variables of interest emerged as significant predictors of R-Span score. In contrast to both hypotheses, results indicate that PSWQ score and estradiol are not significantly related to R-Span Score.

O-Span task.

Results of the full model for O-Span score are presented in Table 9 and estimates for the intercept, estradiol, PSWQ score and their interaction are depicted in Table B11. No variables of interest emerged as significant predictors of O-Span score. Results do not support that O-Span scores are related to PSWQ score and estradiol.

DISCUSSION

To better elucidate the relationship between worry and working memory, the current study investigated relationships between worry and behavioral and neurophysiological measures of working memory across multiple tasks while also taking ovarian hormones into account in women across the menstrual cycle. On the *N*-back task, it was hypothesized that worry would relate to deficits in WM performance (either in longer RT or reduced accuracy) and a reduced P300 amplitude when the task was most difficult (i.e. at higher loads and more difficult trial types of the N-back), but only when estradiol was high. For complex span tasks, it was expected that, if worry related to task-general WM, an increase in worry would be related to poorer task performance on both span tasks when estradiol was high. If worry related to phonologoical/verbal task-specific working memory, deficits related to worry were expected to occur on the R-Span task only when estradiol was high. Results revealed mixed evidence for study hypotheses. Specifically, results indicate that increases in worry symptoms across the menstrual cycle were related to reduced accuracy on 3-back trials of the *N*-back when estradiol levels were low. In contrast to other hypotheses, worry and estradiol were not found to be related to *N*-back reaction time, P300 amplitude, O-Span score, and R-Span score. Findings suggest that worry's relationship with WM may be specific to accuracy on more difficult trials on the *N*-back task when a women's estradiol is relatively low.

The marginally significant interaction between PSWQ score, estradiol and trial type provided partial support for my hypotheses. Greater worry did relate to poorer task performance when the task was more difficult. However, contrary to my hypotheses, this effect occurred only when estradiol was relatively low for the participant. Although this finding is not in line with work by Shansky and colleagues, it may be supported by the anxiety literature, which has

indicated estradiol can have protective effects on other relevant functions such as fear extinction (e.g. Li & Graham, 2017; Montoya & Bos, 2017).

With regards to the null findings for the relationships between worry and other measures of working memory, several possibilities must be considered. First, this study is highly novel in its methodology. Most studies of anxiety and working memory are cross-sectional in design whereas the current study took a repeated measures longitudinal approach. Moreover, previous studies of the role of ovarian hormones in cognition and anxiety typically do not obtain daily assays of hormones from naturally-cycling women across many time points. It is also important to note that the models utilized in the analysis are more complex than analyses typically utilized in cross-sectional work. A larger sample will be critical for providing the opportunity for more well powered tests of what appear to be more complex associations between worry and working memory measures. This study aims to gather 160 participants with useable data, so power concerns will be reduced, and, thus, more robust tests possible in future work.

Although increasing power is important for future analyses, other considerations are equally crucial for improving our understanding of the relationships between worry and working memory. First, the current analyses focused on associations between within-person changes in variables of interest. However, mean levels across all the visit days may also be related to WM measures, as this would reveal between-person effects – e.g., an individual with higher average worry might have lower working memory than a person who has lower worry on average. Future analyses could combine within- and between- person analyses to further address these different effects. Recent studies have also indicated that estradiol may have a quadratic relationship with working memory – taking the form of an "inverted U" (Jacobs & D'Esposito, 2011; Rosenberg & Park, 2002). Future analyses should attempt to examine the effect of a quadratic estradiol term

in the current model once adequately powered to do so. Finally, EEG frequency and advanced signal processing techniques could be considered in future research, especially given the current findings failed to reveal effects on the P300. For example, theta-gamma coupling has recently been presented as a measure of ordering during working memory during the *N*-back task (Lisman & Buzsáki, 2008; Rajji et al., 2016) and may provide a more precise measure of dynamic working memory function compared to time domain ERP measures like the P300.

The role of estradiol in the relationship between worry and working memory was of primary interest in current study. However, the present results have important implications for utilizing the *N*-back task and span tasks for examining working memory function in participants across time. An unstructured covariance structure was found to be necessary to model *N*-back performance and P300 amplitude through examination of the data, as well as through comparison of the selected models with more stringent assumptions of homogeneity of variance and/or covariances (i.e. compound symmetry and heterogeneous compound symmetry structures). The necessity of this structure, along with the great heterogeneity in both variances and covariances observed in estimates generated for the model, suggests that there is important variability related to trial type-by-load interactions that should be accounted for when the *N*-back task is completed by participants multiple times. Future research using the *N*-back in a repeated-measures fashion should examine residuals for trial type-by-load interactions to determine if variances and covariances are heterogeneous, as accounting for such heterogeneity if present is critical for fitting appropriate models of this task.

Related to the more complex structure required to model the *N*-back in the current analysis, the *N*-back presents an additional challenge for longitudinal work through issues of reliability. Traditionally, the N-back has been utilized in neuroimaging work, and the literature

base evaluating its psychometric properties is small and mixed (Jaeggi, Buschkuehl, Perrig, & Meier, 2010). Further examination of the reliability of the task utilized in the current study could lend to the literature on this commonly used working memory measure.

Additionally, analyses for *N*-back reaction time, *N*-back P300 amplitude, O-Span Score, and R-Span score all indicated significant patterning by the order of EEG visits suggesting that there were, indeed, practice effects across all tasks. Such practice effects have been observed in a recent meta-analysis of working memory tasks, even when alternative forms of the tasks were utilized (Scharfen, Jansen, & Holling, 2018). As noted by Scharfen et al. (2018), these practice effects are likely explained by alterations in testing factors unrelated to the task, such as unfamiliarity with task rules or testing anxiety, and the development of task-specific testing strategies and/or memorization of test-specific content. Importantly, it is possible that increases are due to actual gains in WM capacity itself (Scharfen et al., 2018). However, it has been argued that this attributes flexibility to underlying cognitive abilities that is unwarranted, given difficulty training such cognitive abilities in intervention studies where improvement in underlying cognitive abilities is explicitly targeted (Lievens, Reeve, & Heggestad, 2007). It is difficult to disentangle the potential cause of practice effects in the current study, as all explanations could apply to the changes across visit demonstrated in my analyses. Given the focus on worry in this study, it is crucial to consider that state worry at the time of testing could be implicated in the practice effect demonstrated, such that greater worry is experienced by subjects at the first visit due to unfamiliar testing environment and tasks. Incorporating measures of state worry at the time of testing could provide insight into this issue, as our daily worry measure may not have captured worry experienced during the tasks themselves.

Nonetheless, the current study represents a significant step forward in better understanding the intricate relationship between worry and working memory by examining multiple working memory measures and taking ovarian hormones into account in women across the menstrual cycle. Because women experience chronic anxiety characterized by worry more frequently than men, further exploring women's experience of anxiety and related cognitive impairments across the cycle is crucial for improving women's cognitive and emotional health. The present analyses reveal a more nuanced relationship between worry and working memory in women and calls for additional focused research studies into specific groups for whom anxiety and worry are particularly prevalent and problematic. In taking such a targeted approach, the specificity of cognitive models of anxiety and corresponding interventions could greatly be improved.

APPENDICES

APPENDIX A: Primary Tables & Figures

Figures

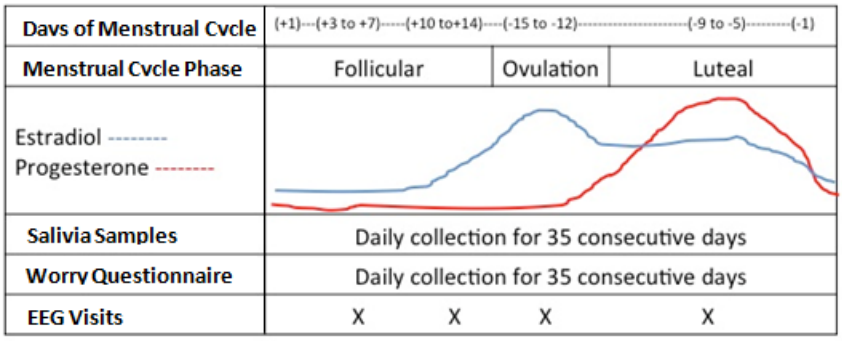


Figure 1. Illustration of study data collection across the menstrual cycle. Day 0 is menstruation. For the purposes of this study, only estradiol levels will be considered. The *N*-back, O-Span, and R-Span tasks are completed at each of the four EEG visits.

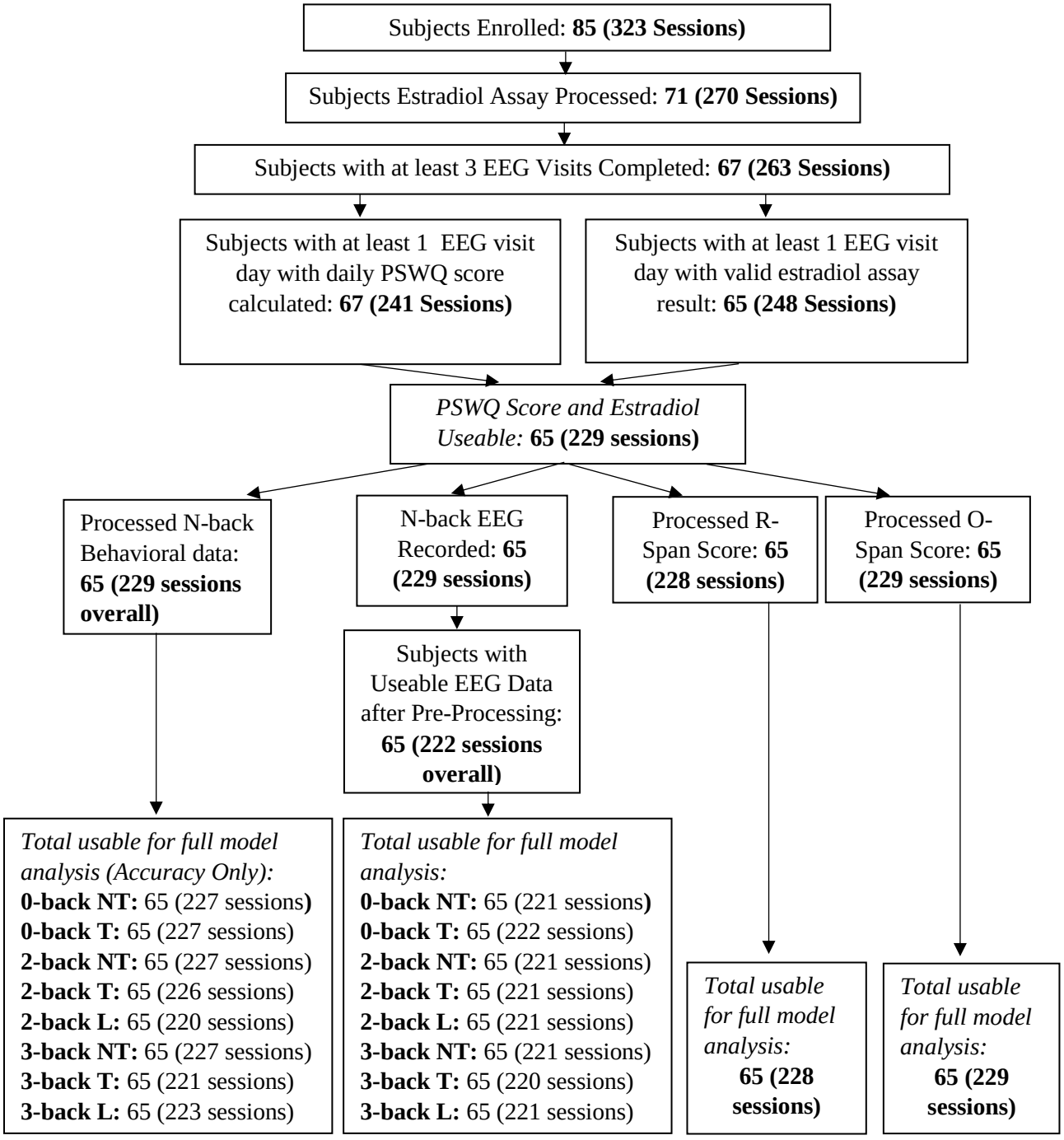
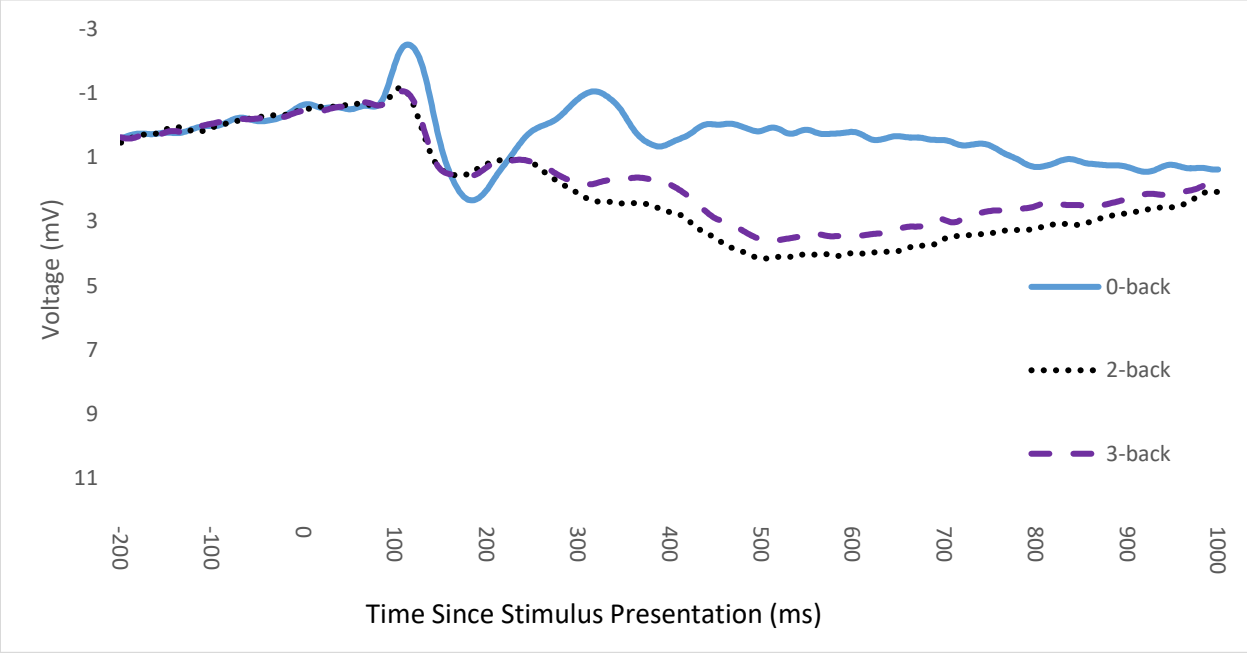
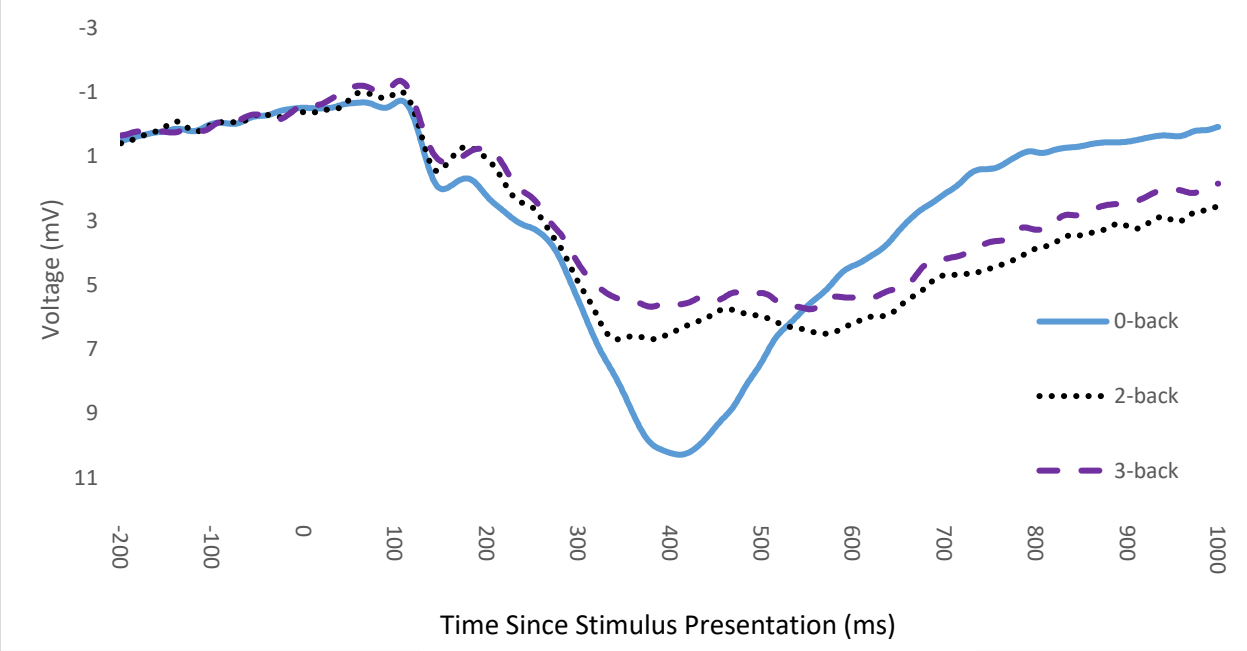


Figure 2. Subject and session loss for each dependent measure of the study.



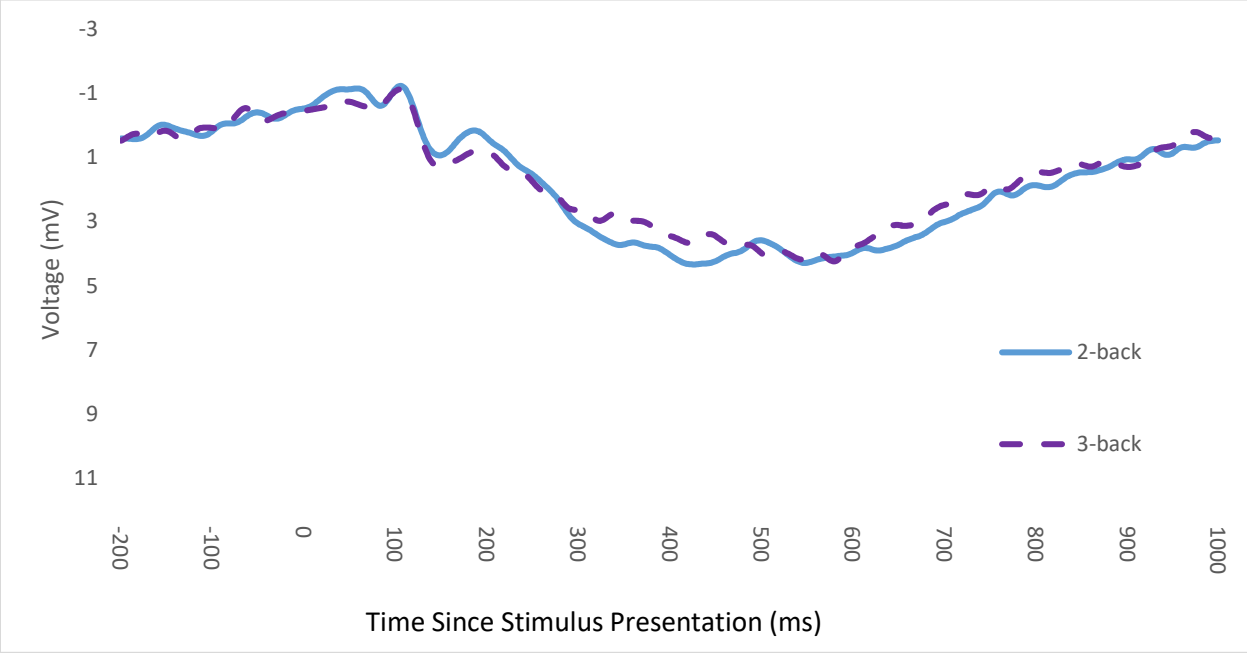
a)



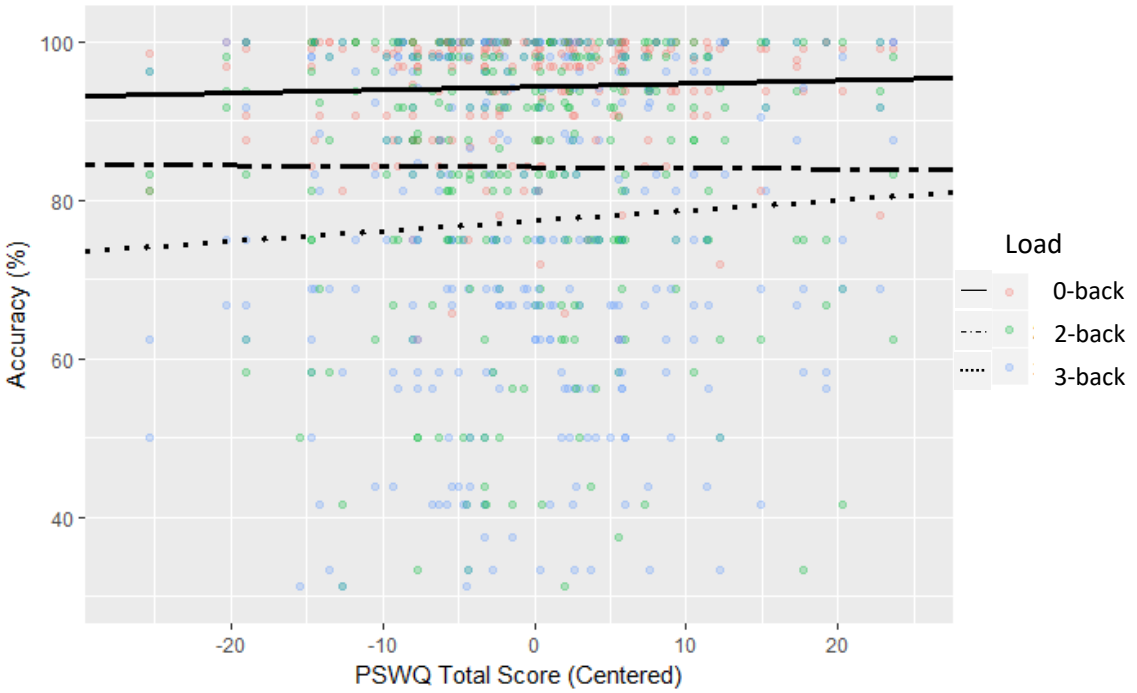
b)

Figure 3. Stimulus-locked grand average waveforms at Pz for **a)** non-targets, **b)** targets, and **c)** lures. Stimulus presentation occurred at 0ms.

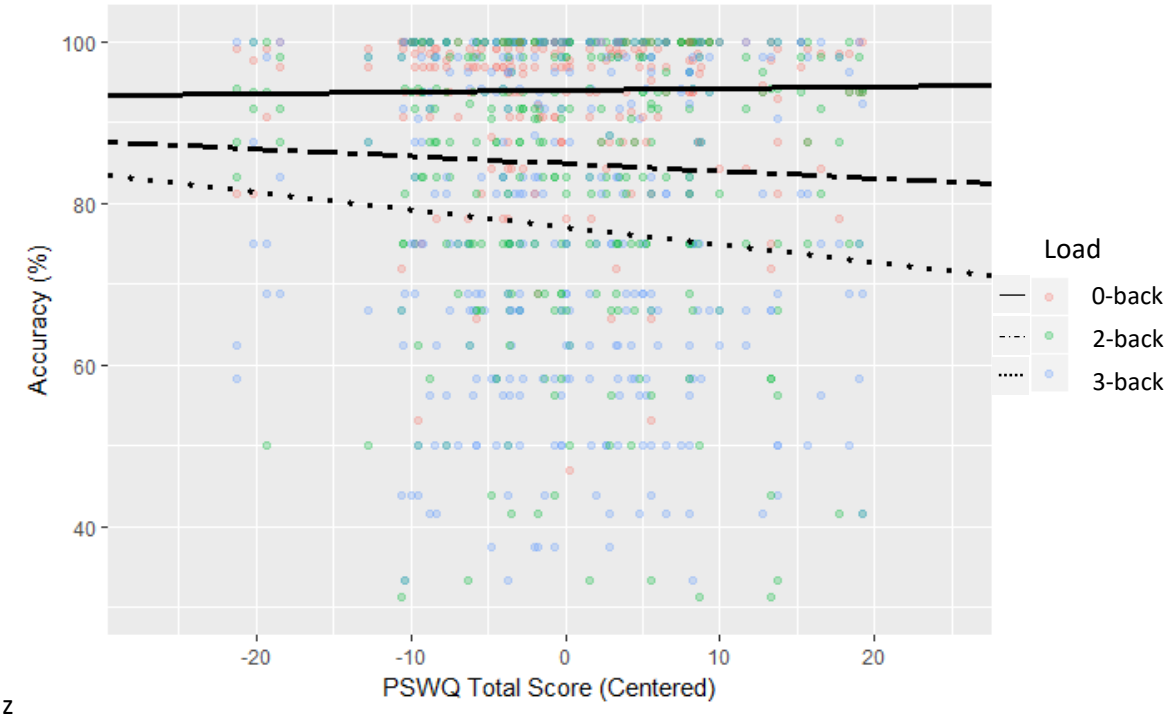
Figure 3 (Cont'd)



c)



a)



z

b)

Figure 4. Breaking down the 3-way PSWQ Score x Estradiol x Load interaction identified in the multilevel model for accuracy. (a) At high levels of estradiol (+1SD), there is no significant

interactions between load and PSWQ Score. (b) At low levels of estradiol (-1SD), there is a PSWQ Score is significantly related to accuracy for 3-back blocks, such that increased worry predicts worse accuracy.

Tables

Table 1: *Demographic Characteristics*

Characteristics	**Statistic**
Age, mean (SD) in years	20.94 (1.81)
Race (%)	
Caucasian/White	65.67
Black/African American	19.40
More than One Race	10.47
Asian	4.48
Hispanic/Latinx (%)	16.42
Sexual Orientation (%)	
Heterosexual	86.57
Bisexual	7.46
Gay/ Lesbian	4.48
Asexual	1.49
Gender Identity	
Female	98.51
Missing	1.49
Income (%)	
$0 - $15,000	29.9
$15,001 - $25,000	14.9
$25,001 - $35,000	3.0
$35,001 - $50,000	4.5
$50,001 - $75,000	9.0
$75,001 - $100,000	10.4
$100,011 - $200,000	20.9
More than $200,000	6.0
Missing	1.5
Financial Supported by Other(s)	61.2

Table 2: *PSWQ score, Estradiol and Span Task Measure Means and Standard Deviations*

Measure	**Mean (SD)**	**Minimum**	**Maximum**
PSWQ Score	41.8 (14.74)	16	79
PSWQ Score-Centered	0 (8.75)	-25.33	23.667
Estradiol (pg/mL)	1.739 (0.779)	0.165	5.478
Estradiol- Centered (pg/mL)	0 (0.53)	-1.8	2.25
R-Span Score	41.79 (18.85)	0	75
O-Span Score	45.36 (19.16)	0	75

Table 3: *N-back Measure Means and Standard Deviations by Load x Trial Type Interaction*

Load x Trial Type Level	**Accuracy (%)**	**Reaction Time (ms)**	**P300 Amp at Pz from 300 – 500ms (mV)**
0-back NT (SD)	98.76 (2.01)	429.42 (74.14)	2.64 (2.76)
0-back T (SD)	89.31 (9.30)	501.06 (70.98)	8.72 (4.70)
2-back NT (SD)	98.04 (2.88)	531.65 (99.59)	2.93 (2.57)
2-back T (SD)	78.64 (16.35)	624.28 (127.85)	6.22 (4.30)
2-back L (SD)	76.38 (18.07)	665.84 (147.61)	3.84 (4.90)
3-back NT (SD)	97.68 (3.09)	543.81 (108.88)	2.28 (2.71)
3-back T (SD)	66.76 (15.34)	664.13 (139.79)	5.34 (4.467)
3-back L (SD)	66.44 (17.33)	640.28 (164.44)	3.25 (4.48)

Table 4: *Test of Type 3 Fixed Effect Significance for Multilevel Model of Accuracy (%) Related to Trial Type, Load, Estradiol, and PSWQ Score Controlling for EEG Visit Number*

Effect	**Num. df**	**Den. df**	***F*-Value**	***p*-value**
Trial Type	2	214	652.48	<.0001
Load	2	219	253.38	<.0001
PSWQ Score	1	195	0.34	0.560
Estradiol	1	198	0.00	0.984
Trial Type x Load	3	215	148.34	<.0001
PSWQ Score x Trial Type	2	215	0.46	0.635
PSWQ Score x Load	2	219	0.25	0.778
Estradiol x Trial Type	2	218	0.14	0.873
Estradiol x Load	2	224	0.91	0.404
PSWQ Score x Estradiol	1	201	2.82	0.095
Estradiol x Load x Trial Type	3	219	0.30	0.824
PSWQ Score x Load x Trial Type	3	215	0.50	0.680
PSWQ Score x Estradiol x Trial Type	2	213	1.29	0.277
PSWQ Score x Estradiol x Load	2	217	2.80	0.063
PSWQ Score x Estradiol x Load x Trial Type	3	212	1.62	0.185
EEG Visit Number	3	154	1.10	0.351

$ICC = 0.007$; $R^2 = 0.723$

Table 5: *Test of Type 3 Fixed Effect Significance for Multilevel Model of Reaction Time Related to Trial Type, Load, Estradiol, and PSWQ Score Controlling for EEG Visit Number*

Effect	**Num. df**	**Den. df**	***F*-Value**	***p*-value**
Trial Type	2	224	379.01	<.0001
Load	2	225	335.21	<.0001
PSWQ Score	1	93.3	0.00	0.963
Estradiol	1	93.9	0.48	0.491
Trial Type x Load	3	224	26.16	<.0001
PSWQ Score x Trial Type	2	224	0.23	0.791
PSWQ Score x Load	2	225	0.69	0.504
Estradiol x Trial Type	2	224	0.65	0.521
Estradiol x Load	2	225	0.01	0.992
PSWQ Score x Estradiol	1	117	0.34	0.560
Estradiol x Load x Trial Type	3	224	0.55	0.648
PSWQ Score x Load x Trial Type	3	224	1.07	0.365
PSWQ Score x Estradiol x Trial Type	2	224	0.34	0.714
PSWQ Score x Estradiol x Load	2	225	2.30	0.103
PSWQ Score x Estradiol x Load x Trial Type	3	224	0.63	0.599
EEG Visit Number	3	158	8.03	<.0001

$ICC = 0.302$, $R^2 = 0.687$

Table 6: *Test of Type 3 Fixed Effect Significance for Multilevel Model of P3 at Pz from 300 to 500ms Related to Trial Type, Load, Estradiol, and PSWQ Score Controlling for EEG Visit Number*

Effect	**Num. df**	**Den. df**	***F*-Value**	***p*-value**
Trial Type	2	218	189.46	<.0001
Load	2	213	54.03	<.0001
PSWQ Score	1	128	0.42	0.520
Estradiol	1	128	1.39	0.240
Trial Type x Load	3	213	38.46	<.0001
PSWQ Score x Trial Type	2	218	0.40	0.673
PSWQ Score x Load	2	213	0.00	0.995
Estradiol x Trial Type	2	217	0.13	0.879
Estradiol x Load	2	212	0.21	0.815
PSWQ Score x Estradiol	1	169	0.20	0.655
Estradiol x Load x Trial Type	3	212	0.15	0.927
PSWQ Score x Load x Trial Type	3	213	0.35	0.791
PSWQ Score x Estradiol x Trial Type	2	217	0.61	0.545
PSWQ Score x Estradiol x Load	2	212	0.14	0.870
PSWQ Score x Estradiol x Load x Trial Type	3	212	0.23	0.874
EEG Visit Number	3	150	13.92	<.0001

ICC = 0.29; R^2 = 0.336

Table 7: *T-Scores for Differences between Least Squares Means for P3 at Pz from 300 to 500ms for Each Load by Trial Type Level*

Load x Trial Type Levels	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1. 0-back non-target							
2. 2-back non-target	-1.94						
3. 3-back non-target	2.46	3.97*					
4. 0-back target	-20.49*	-22.84*	-20.65*				
5. 2-back target	-12.37*	-12.86*	-13.82*	7.58*			
6. 3-back target	-8.35*	-9.00*	-9.81*	11.46*	2.98		
7. 2-back lure	-4.08*	-2.89	-5.19*	11.33*	6.90*	3.65*	
8. 3-back lure	-1.97	-1.24	-3.18*	17.96*	8.84*	6.86*	1.45

Table 8: *Test of Type 3 Fixed Effect Significance for Multilevel Model of R-Span Score related to Estradiol, and PSWQ Score Controlling for EEG Visit Number*

Effect	**Num. df**	**Den. df**	***F*-Value**	***p*-value**
PSWQ Score	1	158	1.80	0.182
Estradiol	1	110	2.55	0.113
PSWQ Score x Estradiol	1	178	1.72	0.192
EEG Visit Number	3	160	3.66	0.014

ICC = 0.74; R^2 = 0.13; Random Intercept Estimate = 255.16; Residual Variance Estimate= 88.68

Table 9: *Test of Type 3 Fixed Effect Significance for Multilevel Model of O-Span Score related to Estradiol and PSWQ Score Controlling for EEG Visit Number*

Effect	**Num. df**	**Den. df**	***F*-Value**	***p*-value**
PSWQ Score	1	159	0.45	0.503
Estradiol	1	159	2.22	0.139
PSWQ Score x Estradiol	1	179	1.46	0.229
EEG Visit Number	3	160	6.76	<0.001

ICC = 0.74; R^2 = 0.06; Random Intercept Estimate = 265.09, Residual variance estimate = 95.20

APPENDIX B: Ancillary Tables

Table B1: *Residual Correlation & Variance Estimates for Multilevel Model of Accuracy (%) Related to Trial Type, Load, Estradiol, and PSWQ Score Controlling for EEG Visit Number*

Load x Trial Type Levels	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1. 0-back non-target	2.86							
2. 0-back target	0.379	93.80						
3. 2-back non-target	0.440	0.297	6.41					
4. 2-back target	0.270	0.338	0.40	286.89				
5. 2-back lure	0.045	0.212	0.191	0.321	332.62			
6. 3-back non-target	0.202	0.285	0.521	0.392	0.209	7.87		
7. 3-back target	0.208	0.289	0.300	0.602	0.266	0.258	242.34	
8. 3-back lure	-0.107	-0.039	0.165	0.0693	0.466	0.120	0.082	296.99

Random intercept estimate for subject = 1.131.

Table B2: *Estimates for the Intercept and Continuous Variable Coefficients from Multilevel Model of Accuracy (%) Related to Trial Type, Load, Estradiol, and PSWQ Score Controlling for EEG Visit Number*

Effect	**Estimate**	**Standard Error**
Intercept	84.132	0.550
PSWQ Score	-0.022	0.062
Estradiol	-0.025	1.018
PSWQ x Estradiol	0.094	0.115

Table B3: *Least Squares Means for Accuracy for Each Load by Trial Type Level*

Load x Trial Type Levels	**Least Squares Mean (%)**	**Standard Error**
0-back non-target	98.752	0.173
0-back target	89.250	0.588
2-back non-target	98.019	0.207
2-back target	78.255	1.029
2-back lure	75.938	1.303
3-back non-target	97.658	0.217
3-back target	66.170	0.985
3-back lure	66.145	1.087

Table B4: *Residual Correlation & Variance Estimates for Multilevel Model of Reaction Time (ms) Related to Trial Type, Load, Estradiol, and PSWQ Score Controlling for EEG Visit Number*

Load x Trial Type Levels	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1. 0-back non-target	1348.8							
2. 0-back target	0.401	2051.3						
3. 2-back non-target	0.270	-0.050	4299.1					
4. 2-back target	0.205	0.226	0.462	11454				
5. 2-back lure	0.120	0.0332	0.562	0.443	15576			
6. 3-back non-target	0.217	-0.090	0.740	0.306	0.209	6424.8		
7. 3-back target	-0.041	0.029	0.425	0.598	0.416	0.466	14143	
8. 3-back lure	-0.026	-0.104	0.546	0.340	0.533	0.647	0.520	19824

Random intercept estimate for subject = 4056.99.

Table B5: *Estimates for the Intercept and Continuous Variable Coefficients from Multilevel Model of Reaction Time (ms) Related to Trial Type, Load, Estradiol, and PSWQ Score Controlling for EEG Visit Number*

Effect	**Estimate**	**Standard Error**
Intercept	579.98	8.956
PSWQ Score	0.172	0.483
Estradiol	-6.306	7.992
PSWQ x Estradiol	-0.783	0.964

Table B6: *Least Squares Means for Reaction Time for Each Load by Trial Type Level*

Load x Trial Type Levels	**Least Squares Mean (ms)**	**Standard Error**
0-back non-target	427.82	8.286
0-back target	498.47	8.468
2-back non-target	531.89	9.029
2-back target	628.26	10.619
2-back lure	673.22	11.435
3-back non-target	545.04	9.529
3-back target	664.27	11.151
3-back lure	645.93	12.219

Table B7: *Residual Correlation & Variance Estimates for Multilevel Model of of P3 at Pz from 300 to 500ms Related to Trial Type, Load, Estradiol, and PSWQ Score Controlling for EEG Visit Number*

Load x Trial Type Levels	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1. 0-back non-target	2.970							
2. 0-back target	-0.266	13.339						
3. 2-back non-target	0.052	0.145	2.689					
4. 2-back target	-0.122	0.128	0.190	14.227				
5. 2-back lure	0.200	-0.241	0.015	0.230	19.975			
6. 3-back non-target	0.230	-0.336	0.038	-0.020	0.179	3.660		
7. 3-back target	-0.333	0.345	0.203	0.360	-0.043	-0.143	15.915	
8. 3-back lure	-0.279	0.309	0.219	0.184	0.009	-0.113	0.366	16.246

Random intercept estimate for subject = 4.539

Table B8: *Estimates for the Intercept and Continuous Variable Coefficients from Multilevel Model of P3 at Pz from 300 to 500ms Related to Trial Type, Load, Estradiol, and PSWQ Score Controlling for EEG Visit Number*

Effect	**Estimate**	**Standard Error**
Intercept	4.31	0.289
PSWQ Score	<0.001	0.014
Estradiol	-0.27	0.221
PSWQ x Estradiol	0.02	0.028

Table B9: *Least Squares Means for P3 at Pz from 300 to 500ms for Each Load by Trial Type Level*

Load x Trial Type Levels	**Least Squares Mean (mV)**	**Standard Error**
0-back non-target	2.49	0.289
0-back target	8.59	0.361
2-back non-target	2.79	0.287
2-back target	6.09	0.367
2-back lure	3.70	0.400
3-back non-target	2.13	0.295
3-back target	5.21	0.377
3-back lure	3.12	0.379

Table B10: *Estimates for the Intercept and Continuous Variable Coefficients from Multilevel Model of R-Span Score related to Estradiol and PSWQ Score Controlling for EEG Visit Number*

Effect	**Estimate**	**Standard Error**
Intercept	41.91	2.086
PSWQ Score	1.64	1.220
Estradiol	-0.03	0.073
PSWQ x Estradiol	0.23	0.174

Table B11: *Estimates for the Intercept and Continuous Variable Coefficients from Multilevel Model of R-Span Score related to Estradiol and PSWQ Score Controlling for EEG Visit Number*

Effect	**Estimate**	**Standard Error**
Intercept	41.91	2.086
PSWQ Score	1.64	1.220
Estradiol	-0.03	0.073
PSWQ x Estradiol	0.23	0.174

REFERENCES

REFERENCES

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Sage.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

Baddeley, A. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences*, *93*(24), 13468–13472.

Baddeley, A. D. (1986). *Working memory* (Vol. 11). Elsevier.

Bailey, K., Mlynarczyk, G., & West, R. (2016). Slow wave activity related to working memory maintenance in the N-back task. *Journal of Psychophysiology*, *30*(4), 141.

Borkovec, T. D., Ray, W. J., & Stober, J. (1998). Worry: A cognitive phenomenon intimately linked to affective, physiological, and interpersonal behavioral processes. *Cognitive Therapy and Research*, *22*(6), 561–576.

Bredemeier, K., & Berenbaum, H. (2013). Cross-sectional and longitudinal relations between working memory performance and worry. *Journal of Experimental Psychopathology*, *4*(4), 420–434.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion*, *6*(6), 409–434.

Eysenck, M. W., & Derakshan, N. (2011). New perspectives in attentional control theory. *Personality and Individual Differences*, *50*(7), 955–960.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion*, *7*(2), 336.

Ganley, C. M., & Vasilyeva, M. (2014). The role of anxiety and working memory in gender differences in mathematics. *Journal of Educational Psychology*, *106*(1), 105.

Gasbarri, A., Pompili, A., d'Onofrio, A., Cifariello, A., Tavares, M. C., & Tomaz, C. (2008). Working memory for emotional facial expressions: role of the estrogen in young women. *Psychoneuroendocrinology*, *33*(7), 964–972.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*(3), 316–322.

Jacobs, E., & D'Esposito, M. (2011). Estrogen Shapes Dopamine-Dependent Cognitive Processes: Implications for Women's Health. *Journal of Neuroscience*, *31*(14), 5286–5293. https://doi.org/10.1523/JNEUROSCI.6394-10.2011

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*(4), 394–412. https://doi.org/10.1080/09658211003702171

Kane, M. J., Conway, A. R., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. *Variation in Working Memory*, *1*, 21–48.

Kappenman, E. S., & Luck, S. J. (2011). ERP Components: The Ups and Downs of Brainwave Recordings. *The Oxford Handbook of Event-Related Potential Components*, 3.

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, H.-U. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *International Journal of Methods in Psychiatric Research*, *21*(3), 169–184.

Klump, K. L., O'Connor, S. M., Hildebrandt, B. A., Keel, P. K., Neale, M., Sisk, C. L., … Burt, S. A. (2016). Differential Effects of Estrogen and Progesterone on Genetic and Environmental Risk for Emotional Eating in Women. *Clinical Psychological Science*, *4*(5), 895–908. https://doi.org/10.1177/2167702616641637

Li, S. H., & Graham, B. M. (2017). Why are women so vulnerable to anxiety, trauma-related and stress-related disorders? The potential role of sex hormones. *The Lancet Psychiatry*, *4*(1), 73–82.

Lievens, F., Reeve, C. L., & Heggestad, E. D. (2007). An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. *Journal of Applied Psychology*, *92*(6), 1672.

Lisman, J., & Buzsáki, G. (2008). A neural coding scheme formed by the combined function of gamma and theta oscillations. *Schizophrenia Bulletin*, *34*(5), 974–980.

Maeng, L. Y., & Milad, M. R. (2015). Sex differences in anxiety disorders: interactions between fear, stress, and gonadal hormones. *Hormones and Behavior*, *76*, 106–117.

Man, M. S., MacMillan, I., Scott, J., & Young, A. H. (1999). Mood, neuropsychological function and cognitions in premenstrual dysphoric disorder. *Psychological Medicine*, *29*(3), 727–733.

McEvoy, L. K., Smith, M. E., & Gevins, A. (1998). Dynamic cortical networks of verbal and spatial working memory: effects of memory load and task practice. *Cerebral Cortex (New York, NY: 1991)*, *8*(7), 563–574.

McLean, C. P., Asnaani, A., Litz, B. T., & Hofmann, S. G. (2011). Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. *Journal of Psychiatric Research*, *45*(8), 1027–1035.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy*, *28*(6), 487–495.

Montoya, E. R., & Bos, P. A. (2017). How oral contraceptives impact social-emotional behavior and brain function. *Trends in Cognitive Sciences*, *21*(2), 125–136.

Moran, T. P. (2016). Anxiety and working memory capacity: A meta-analysis and narrative review. *Psychological Bulletin*, *142*(8), 831–864. http://dx.doi.org.proxy2.cl.msu.edu/10.1037/bul0000051

Nitschke, J. B., Heller, W., Imig, J. C., McDonald, R. P., & Miller, G. A. (2001). Distinguishing Dimensions of Anxiety and Depression. *Cognitive Therapy and Research*, *25*(1), 1–22. https://doi.org/10.1023/A:1026485530405

Polich, J. (2012). Neuropsychology of P300. *Oxford Handbook of Event-Related Potential Components*, 159–188.

Rajji, T. K., Zomorrodi, R., Barr, M. S., Blumberger, D. M., Mulsant, B. H., & Daskalakis, Z. J. (2016). Ordering information in working memory and modulation of gamma by theta oscillations in humans. *Cerebral Cortex*, *27*(2), 1482–1490.

Rosenberg, L., & Park, S. (2002). Verbal and spatial functions across the menstrual cycle in healthy young women. *Psychoneuroendocrinology*, *27*(7), 835–841.

Scharfen, J., Jansen, K., & Holling, H. (2018). Retest effects in working memory capacity tests: A meta-analysis. *Psychonomic Bulletin & Review*, 1–25.

Schoofs, D., Pabst, S., Brand, M., & Wolf, O. T. (2013). Working memory is differentially affected by stress in men and women. *Behavioural Brain Research*, *241*, 144–153.

Shansky, R. M., Glavis-Bloom, C., Lerman, D., McRae, P., Benson, C., Miller, K., … Arnsten, A. F. T. (2004). Estrogen mediates sex differences in stress-induced prefrontal cortex dysfunction. *Molecular Psychiatry*, *9*(5), 531.

Shansky, R., Rubinow, K., Brennan, A., & Arnsten, A. F. (2006). The effects of sex and hormonal status on restraint-stress-induced working memory impairment. *Behavioral and Brain Functions*, *2*(1), 8.

Shansky, Rebecca M., Bender, G., & Arnsten, A. F. T. (2009). Estrogen prevents norepinephrine alpha-2a receptor reversal of stress-induced working memory impairment. *Stress*, *12*(5), 457–463.

Shansky, Rebecca M., & Lipps, J. (2013). Stress-induced cognitive dysfunction: hormone-neurotransmitter interactions in the prefrontal cortex. *Frontiers in Human Neuroscience*, *7*.

Shields, G. S., Sazma, M. A., & Yonelinas, A. P. (2016). The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. *Neuroscience & Biobehavioral Reviews*, *68*, 651–668. https://doi.org/10.1016/j.neubiorev.2016.06.038

Stefanopoulou, E., Hirsch, C. R., Hayes, S., Adlam, A., & Coker, S. (2014). Are Attentional Control Resources Reduced by Worry in Generalized Anxiety Disorder? *Journal of Abnormal Psychology*, *123*(2), 330–335. https://doi.org/10.1037/a0036343

Trezise, K., & Reeve, R. A. (2014). Working memory, worry, and algebraic ability. *Journal of Experimental Child Psychology*, *121*, 120–136.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.

Watter, S., Geffen, G. M., & Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology*, *38*(6), 998–1003. https://doi.org/10.1111/1469-8986.3860998

West, R., Bowry, R., & Krompinger, J. (2006). The effects of working memory demands on the neural correlates of prospective memory. *Neuropsychologia*, *44*(2), 197–207.